

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-285085

(43)Date of publication of application : 13.10.2000

(51)Int.Cl. G06F 15/177
G06F 15/16

(21)Application number : 2000-034267

(71)Applicant : LUCENT TECHNOL INC

(22)Date of filing : 10.02.2000

(72)Inventor : THOMAS J EDWARDS
DAVID MARLEN

(30)Priority

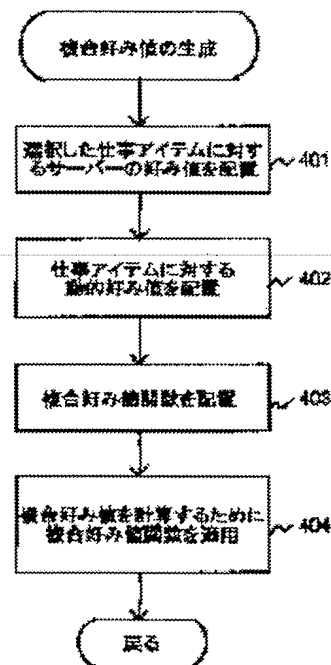
Priority number : 99 247893 Priority date : 10.02.1999 Priority country : US

(54) METHOD FOR DECIDING WHETHER SERVER SHOULD BE ASSIGNED TO SERVER LOOP SET TO WORK TYPE IN WORK PROCESSING FACILITY

(57)Abstract:

PROBLEM TO BE SOLVED: To secure the work activity for the best work item with priority on the hierarchy by adjusting the number of servers available to specific work types in response to the service needs which are not satisfied with the work processing facilities or the service needs kept in the crisis states.

SOLUTION: The liking value of servers are arranged against the work type of a selected item (401), and the dynamic liking value are arranged according to the service needs kept in the crisis states such as the service level acquired to a work item, the queue state, the desired service level, the warning state, the response delay, the call disuse rate, the intervention of an operator, etc., (402). Then a compound liking value function that is currently selected is set in the work facilities (403) and applied for calculating the compound liking value (404). Thus, the number of available servers is adjusted against specific work types.



LEGAL STATUS

[Date of request for examination] 31.05.2001

[Date of sending the examiner's decision of rejection] 16.03.2005

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3844932

[Date of registration] 25.08.2006

[Number of appeal against examiner's decision of rejection] 2005-11073
[Date of requesting appeal against examiner's decision of rejection] 14.06.2005
[Date of extinction of right]

(11)特許出願公開番号

特開2000-285085

(P2000-285085A)

(43)公開日 平成12年10月13日(2000.10.13)

(51) Int.Cl. ⁷	識別記号	F I	テ-マ-コ-ト*(参考)
G 0 6 F 15/177	6 7 4	G 0 6 F 15/177	6 7 4 A
15/16	6 2 0	15/16	6 2 0 B

審査請求 未請求 請求項の数10 OL 外国語出願 (全 70 頁)

(21)出願番号 特願2000-34267(P2000-34267)

(22) 出題日 平成12年2月10日(2000.2.10)

(31)優先権主張番号 09/247893

(32)優先日 平成11年2月10日(1999.2.10)

(33)優先権主張国 米国 (US)

(71)出題人 596077259

ルーセント テクノロジーズ インコーポ
レイテッド

Lucent Technologies
Inc.

アメリカ合衆国 07974 ニュージャージー
ー、マレーヒル、マウンテン アベニュー
600-700

(74) 代理人 100081053

弁理士 三俣 弘文

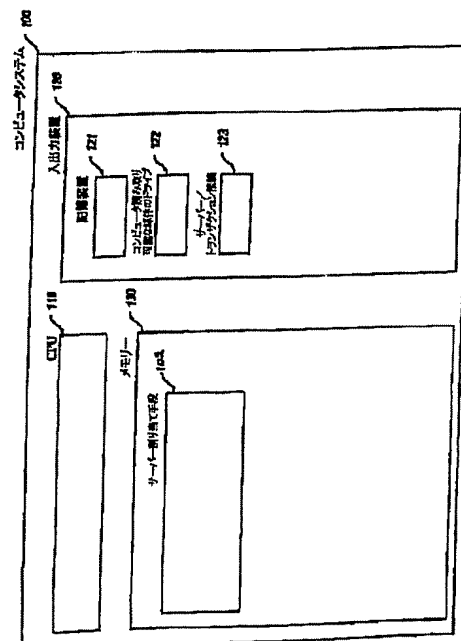
[最終頁に続く](#)

(54) 【発明の名称】 仕事処理設備にてサーバーを仕事種類に対するサーバープールに割り当てるべきかを判断する方法

(57) 【要約】

【課題】 サービスの提供における資源の割り当てを改善する。

【解決手段】 サービスレベル目標の達成を最大化するために、仕事処理設備において待ち行列に入れられた仕事アイテムに関する満足されていないサービス要求に対して階層的調整を提供する。階層的調整手段は仕事分配手段、自動呼び配信（ACD）、あるいは同様な設備にて用いることができる。この階層的調整手段は通常、サーバーが仕事をするため可用となり、複数の仕事種類の仕事アイテムが待ち行列に入れられ分配を待つような場合にそのタスクを実行する。この階層的調整手段は、待ち行列に入った仕事アイテムのどの仕事アイテムが可用サーバーに分配されるかを判定する。この階層的調整手段は、仕事アイテムに対する複合好み値に基づいて他の待ち行列に入れられた仕事アイテムの中から仕事アイテムを選択することができる。



【特許請求の範囲】

【請求項 1】 仕事処理設備にてサーバーを仕事種類に対するサーバープールに割り当てるべきかを判断する方法であって、前記サーバープールは、仕事種類の仕事を処理するサーバーの指示を保持するように構成し、

(A) サーバーの仕事種類に対する第 1 の値がサーバーの仕事種類に対する第 2 の値よりも小さいかどうかを判断するステップと、ここで、第 1 の値はその仕事種類に対するサーバーのポジティブ度合いを示し、第 2 の値はその仕事種類に対するサーバーのネガティブ度合いを示し、

(B) サーバーの仕事種類に対する第 1 の値がサーバーの仕事種類に対する第 2 の値よりも小さければ、その仕事種類に対する複合好み値を判断するステップと、

(C) 仕事種類に対する複合好み値がサーバーの仕事種類に対する第 2 の値よりも大きいかどうかを判断するステップと、

(D) 仕事種類に対する複合好み値がサーバーの仕事種類に対する第 2 の値より大きければ、仕事処理設備に対し、その仕事種類に対するサーバープールへと割り当てることに当該サーバーが適していることの指示を送るステップとを有することを特徴とする方法。

【請求項 2】 前記ステップ (B) は、

(a) 当該仕事種類に対して動的好み値を決定するステップと、ここで、決定した動的好み値は、当該仕事種類の仕事をサーバーに行わせる仕事処理設備の必要性の尺度を繁栄し、

(b) 複合好み値を決定するように構成するユーザーが選択可能な複合好み値関数にアクセスするステップと、

(c) 当該仕事種類に対する複合好み値を生成するためにアクセスしたユーザーが選択可能な好み値関数に当該仕事種類に判断したサーバーの好み値および当該仕事種類に対する判断した動的好み値を供給するステップとを有することを特徴とする請求項 1 記載の方法。

【請求項 3】 (E) 当該仕事種類に対する複合好み値がサーバーの仕事種類に対する第 2 値より大きくない限り、サーバーの仕事種類に対する第 1 の値がサーバの仕事種類に対する第 2 の値よりも小さければ、当該仕事種類のサーバープールに対してサーバーを割り当てないステップを更に有することを特徴とする請求項 1 記載の方法。

【請求項 4】 (F) 当該仕事種類に対する複合好み値を再度判断するステップと、

(G) 仕事種類に対する再度判断した複合好み値がサーバーの第 2 の値よりも大きいかどうかを判断するステップと、

(H) 再度判断した複合好み値がサーバーの第 2 の値よりも小さければ、仕事処理設備に対し、当該仕事種類に対するサーバープールからサーバーを除くべきであることの更なる指示を送るステップとを特徴とする請求項

1 記載の方法。

【請求項 5】 当該仕事処理設備は、複数の仕事種類に対する複数のサーバープールを有し、仕事種類に対するサーバーの第 1 の値が仕事種類に対するサーバーの第 2 の値より小さければ、複数の仕事種類の各仕事種類に対して複数の好み値が判断されることを特徴とする請求項 1 記載の方法。

【請求項 6】 仕事処理設備にてサーバーを仕事種類に対するサーバープールに割り当てるべきかを判断する方法であって、前記サーバープールは、仕事種類の仕事を処理するサーバーの指示を保持するように構成し、

(A) サーバーの仕事種類に対する第 1 の値がサーバーの仕事種類に対する第 2 の値よりも小さいかどうかを判断するように構成する第 1 比較手段と、第 1 の値はその仕事種類に対するサーバーのポジティブ度合いを示し、第 2 の値はその仕事種類に対するサーバーのネガティブ度合いを示し、

(B) サーバーの仕事種類に対する第 1 の値がサーバーの仕事種類に対する第 2 の値よりも小さければ、その仕事種類に対する複合好み値を判断するように構成する第 2 比較手段と、

(C) 仕事種類に対する複合好み値がサーバーの仕事種類に対する第 2 の値よりも大きいかどうかを判断するように構成する第 3 比較手段と、

(D) 仕事種類に対する複合好み値がサーバーの仕事種類に対する第 2 の値より大きければ、仕事処理設備に対し、その仕事種類に対するサーバープールへと割り当てることに当該サーバーが適していることの指示を送るように構成する結果指示手段とを有することを特徴とするシステム。

【請求項 7】 第 2 比較手段 (B) は、

(a) 当該仕事種類に対して動的好み値を判断するように構成する判断手段と、ここで、決定した動的好み値は、当該仕事種類の仕事をサーバーに行わせる仕事処理設備の必要性の尺度を繁栄し、

(b) 当該仕事種類に対する複合好み値を生成するためにアクセスしたユーザーが選択可能な好み値関数に当該仕事種類に判断したサーバーの好み値および当該仕事種類に対する判断した動的好み値を供給するように構成するプロセッサとを有することを特徴とする請求項 6 記載のシステム。

【請求項 8】 (E) 前記結果指示手段 (D) は、当該仕事種類に対する複合好み値がサーバーの仕事種類に対する第 2 値より大きくない限り、サーバーの仕事種類に対する第 1 の値がサーバの仕事種類に対する第 2 の値よりも小さければ、当該仕事種類のサーバープールに対してサーバーを割り当てないように指示するように構成することを特徴とする請求項 6 記載のシステム。

【請求項 9】 (F) 第 2 比較手段は、当該仕事種類に対する複合好み値を再度判断するように構成し、

(G) 仕事種類に対する再度判断した複合好み値がサーバーの第2の値よりも大きいかどうかを判断するように構成する第4比較手段と、

(H) 再度判断した複合好み値がサーバーの第2の値よりも小さければ、仕事処理設備に対し、当該仕事種類に対するサーバープールからサーバーを除くべきであることの更なる指示を送るように構成する第2結果指示手段とを更に有することを特徴とする請求項6記載のシステム。

【請求項10】 当該仕事処理設備は、複数の仕事種類に対する複数のサーバープールを有し、第2比較手段は、仕事種類に対するサーバーの第1の値が仕事種類に対するサーバーの第2の値より小さければ、複数の仕事種類の各仕事種類に対して複数の好み値を判断するように更に構成することを特徴とする請求項6記載のシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、サービスの提供における資源の割り当てに関する。

【0002】

【従来の技術】 目的を達成するために多くのサービス組織はサーバー（例えば、サービスを提供するオペレーター）を動的に割り当てる必要がある。このような割り当ては通常、マニュアル的に行われる。サーバーには、人間のおよび非人間的両方のサービスエージェントが含まれる。サービスの外的パフォーマンスの尺度は益々、利用度や労働コストのような内的コスト尺度に対し大きくなっている。このような外的尺度はしばしば、所望の目的を満たすか満たさないかと特定のトランザクションを分類し、このような目的を満たすトランザクションの割合を判定することを含んでいる。このような割合はサービスレベルと呼ばれる。サービスレベルは特定の時間にわたって、あるいは特定のトランザクションの数にわたって測定される。

【0003】 サービスレベルの例として、更にアクションをせずに解決した顧客問題の割合、約束した時間内でタクシーが顧客の場所へと到着した割合、別のサーバーへと転送したり別のサーバーを参照したりすることを必要とせずに品質を保って代表へと取り扱われる電話呼びの割合、遅れずにサーバーへと接続することができる電話呼びの割合、24時間以内に応答される電子メール要求の割合、平日で特定のバスルートでの市内バスの時間通りの出発の割合、顧客の苦情を発生せずに取り扱われたトランザクションの割合、フル品質のサーバーによって優先顧客の呼びが取り扱われた割合、スペイン語を話せるサーバーによってスペイン系の顧客が取り扱われた割合、サーバーに接続する前に顧客が電話呼びをあきらめなかった割合、中央局スイッチにてブロックされなかった顧客の問い合わせ電話呼びの割合、表示されるまで

あきらめなかったセルフサービスのウェブページでの顧客セッションの割合、話し中にその場で電話で完了することができる顧客要求の割合、要求から1時間半以内で処理できたローン申請の割合、品質が高いサーバーで適切に取り扱い8秒以内で応答した有線電話呼びの割合等がある。

【0004】 このコンテキストにおけるサービス組織のサービスレベルの目標は、サービスレベルの特定の望ましい値である。その目標は、達成したサービスレベルがその目標に対する所望のサービスレベル以上であれば満足したものとされる。逆に、反対の場合は満足されなかったとされる。例えば、もしその日の優先顧客から電話呼びのうち87%が12秒以内で応答された場合、その日の優先顧客からの電話呼びの85%以上を12秒以内で応答するという目標は達成されたこととなる。逆に、もし84%のみが12秒以内で応答された場合ではこの目標は達成されないものとされる。このフレームワークでは、目標は達成されたかされなかったかのいずれかである。また、目標を越えて非常に高いレベルのサービスに対しては付加的な利益を発生させていない。

【0005】 サービスレベル目標は、例外的な資源の割合を含むコンティンジェンシアクション（contingency action）の基準に一般に用いられている目標である。この事に関連して他のことに関連するパフォーマンス目標には、平均待ち時間最大値、現在の待ち行列状態、仕事シフトにおけるトランザクションスループットの最小数、品質サービスを示す多くの決定可能なファクターを伴う尺度を含む。

【0006】 サービスタイプに対して割り当てられたサーバー資源の数はしばしばそのサービスタイプで達成されるサービスレベルに影響を与える。このような場合、サービスレベルの目標を達成するためにサーバーを再割り当てすることがよくある。しかし、一般にこのような再割り当ては機会のコストを発生させる。なぜなら、他の仕事のサービスレベルに対して悪影響を与えるからである。これは企業の運用規則で確立することができるように適切な優先度階層に基づいてこの機会のコストを適正化することがよくある。

【0007】 例えば、コールセンターのサーバーがローンサービスとセールスサービスの両方のトランザクションを取り扱うことができるものと想定する。セールス活動に多くのサーバーが割り当てられている場合、応答遅延に関してセールスサービストランザクションは高いサービスレベルを達成することができる。すなわち、セールスの呼びそれぞれに応答する時間が減る。同時に、ローンサービスの呼びには余り速く応答されず、ローンサービストランザクションのサービスレベルを減らしてしまう。現在の顧客がローン対称の会社へと切り替わることはあまりなく早く対応しなければ競合他社に仕事を取られてしまうような新しい顧客を獲得する必要があるこ

とのような理由によってローンサービスがセルフサービスよりも重要では無いとしてサービス組織はこれを合理化する。

【0008】このサービス組織はローンサービスの目標を達成したいがセールスサービスの目標を未達成となっても達成したいとは思わない。セールスの目標が危機でなく、ローンサービスの目標が達成されていなければ、サービス組織はローンサービスの資源を増やすことを望む。このサービス組織は両方の目標を達成したいがセールスの目標はローンサービスの目標よりも重要でありそれに先取り (preempt) することができる。すなわち、もし一方の目標のみを満足することができる状態であればセールスの目標を満足させる。

【0009】幾つかのサービス目標のうち、短い「機会ウィンドー」を有する仕事に関連する目標、そして、長い機会ウィンドーを有する仕事に対する目標があることがある。短い機会ウィンドーの例としては、電話の呼びがある。これは、数十秒の間に応答されなければ、発呼者によってあきらめられる。長い機会ウィンドー仕事の例としては、次の徴収サイクルまでに終了しなければならぬ徴収調整に関連する顧客からの手紙がある。このように、短い機会ウィンドー仕事に対してサービス目標を満足させるために非常に短い時間しか与えられず、長い機会ウィンドー仕事に対してサービス目標を達成するための機会ウィンドーはもっと広い。従って、短い機会ウィンドー仕事のサービス目標を満足させるために、長い機会ウィンドー仕事を行う資源を一時的により多く割り当てるような生産的な運用上の戦略をすることができる。

【0010】多くのサービス組織は、所望の結果を得るためにサーバーを動的に割り当てる必要がある。通常このような再割り当てを行うのにマニュアル的な介入がなされる。特定の仕事種類に必要なスキルプロファイルを満足するにも関わらず、その特定の仕事種類で利用可能状態でないようなアベイアンス (abeyance) 状態にサーバーがされていることが多い。しかし、条件が揃えばその仕事種類に割り当てられることがある。このような場合、このサーバーはその仕事に関してバックアップサーバーと呼ばれ、そのサーバーは通常はその仕事種類に対して利用されない。

【0011】ある活動に対してサーバー資源の割り当てを増やすことには通常、他の活動に影響を与える。これら他の活動のそれぞれは潜在的に多くのサービスレベルと関連づけられており、それぞれは達成目標や達成レベルを有する。資源の再割り当ては各資源に対する全ての他の仕事に対してのサービス尺度に依存することができる。このような潜在的複雑な割り当て機能をマニュアル的に行うことは非常に最適から外れた結果を発生させてしまう。マニュアル的割り当ては遅すぎるが多く、再割り当てされたサーバーがそれらの好ましい仕事にす

ぐ戻れないならば多くの問題を発生させてしまう。

【0012】多くの自動呼びディストリビューター (ACD) は、「呼びオーバーフロー」と呼ばれる特徴を有する。呼びオーバーフローは、選択された状態の場合に呼び待ち行列群に対してサーバーグループを利用可能にする。しかし、ACDは一般に、適切な動的状態の場合に呼び群を受けることに対しサーバーをアベイアンス状態に維持するような設備に掛ける。ACDは一般に、サーバーへと呼びを受けるプロセスを明示的に制御することを必要とするが、呼びに対しサーバーを獲得するプロセスにたいしては余り明示的な制御を提供しない。しかし、仕事種類に対してサーバーの観点は仕事の完了を最大化するために他の仕事との間で選択するのに際して重大なことが多い。

【0013】サーバーを割り当てる際に個々のサーバーのスキルを考慮するようなスキルベースのルーティングの出現によってこのような状況は複雑になっている。スキルベースのルーティングは、資源の断片を、スキルの区別が無視されるようなモノリシックな受け皿へと単純化することを許容しない。このように、従来のACD、ワークフローオートメーションシステムはこの条件を満足しない。

【0014】仕事分配システムは、ユーザーに対して、上述のようにサーバーの再割り当てを行う為にサーバーのスキルを操作することを強いる。ACD、他の自動仕事ディストリビューターは、スキル需要、あるいはディストリビューションマップ上の中の種類の待ち行列ポイントに対してのみ「サービスレベル」をレポートする。しかし、スキルがサーバーの能力の属性の変わりに実際に仕事種類であることがある。これら従来の解決方法はサービス組織に重要なサービス組織に対するモニタリングや制御に対し潜在的に厳しい制限を与えてしまう。

【0015】

【発明が解決しようとする課題】この環境において、サービス組織はサービス目標およびこれら目標の規定された優先度に関して最もよい仕事アイテムがサーバーに与えられることを確実にするように、階層上にて仕事活動に優先的取り扱いを与えることを望む。仕事処理設備において仕事種類に関するサービス目標の達成または非達成に基づいて仕事種類に対して利用可能なサーバーの器を動的に広げるような自動化システムの必要性がある。

【0016】

【課題を解決するための手段】本発明は、仕事処理設備において満足されないまたは危機状態のサービスニーズに基づいて特定の仕事種類に対して利用可能なサーバーの数を調整するような設備を提供する。サーバーには、人間あるいは人間ではないサービスエージェントが含まれることができる。サーバー割り当て、対応するサーバー割り当て方法は、サーバーの器に対してのサーバーの割り当てを条件的に調整するように仕事ディストリビューター

一において利用する。仕事アイテムが到来する電話の呼びであるような一般的な仕事ディストリビューターの例としては、自動呼びディストリビューター（ACD）がある。

【0017】一実施例に従うと、サーバーの割り当てにおいて、仕事種類に対する複合好み値を、その仕事種類に対する各サーバーのしきい値と比較する。複合好み値がサーバーのしきい値以上であると判断すると、その仕事種類に対するサーバープール（サーバーの器）にサーバーが含まれることを指示する。各サーバーは異なる仕事種類に対して好み値およびしきい値を有する。好み値の大きさは、その仕事種類に対する親和性を現す。サーバーのしきい値は、その仕事種類に対して仕事を行うことの不本意な度合いを現す。（なお本明細書において、ポジティブ度合いとは好ましい度合いを意味し、ネガティブ度合いとは好ましくない度合いを意味する。）一実施例に従うと、サーバーは通常、好み値がしきい値よりも小さいような仕事種類の仕事を受けないが、サーバーの複合好み値が仕事種類に対するしきい値を越えた場合にはこの限りではない。

【0018】一実施例に従うと、サーバーの割り当てにおいて、多くのユーザが選択可能な複合好み値を計算し、多くのユーザが選択可能な関数を用いる。複合好み値には、サービスレベル、待ち行列状態、所望サービスレベル、警告状態、応答遅延、呼び放棄率、オペレータ介入のような入力から計算される動的好み値が含まれる。一実施例に従うと、複合好み値関数の入力は、仕事種類に対するサーバーの好み値および仕事種類に対する動的好み値を含む。従って、この実施例において、複合好み値は、多くの仕事種類に対しサーバーの特定の好みおよび多くの仕事種類を取り扱うことに対する仕事種類設備の現在の一般的な好みの両方を考慮する。

【0019】サーバー割り当てにおいて、一実施例に従い、仕事種類の動的好み値および仕事種類に対するサーバーの好み値を入力として受ける多くの複合好み値関数の何れをも用いることができる。単純な複合好み値関数として仕事種類に対するサーバーの好み値および仕事種類の動的好み値を合計するものを含む。

【0020】サーバー割り当てにおいて、満足されない目標を有する仕事種類に対するサーバープールにおける拡張実行するために仕事の動的な好都合な需要を適用する。サーバー割り当てにおいて、好都合な仕事に対する運用規則の効果と釣り合うバックアップサーバーの利用を可能にする。また、仕事処理設備やサーバー組織に対して予期しない制約無しでバックアップサーバーの利用を可能にする。

【0021】

【発明の実施の形態】本発明は、サービス目標の達成を最大化するために、仕事処理設備において満足されないサービスニーズに基づき特定の仕事種類に対して利用可

能なサーバーのプールを調整する設備を提供する。サーバーには、人間および人間ではないサービスエージェントが含まれる。サーバー割り当てにおいて、仕事分配手段にて用いられサーバーのプールへのサーバーの割り当てを条件的に調整する。サーバープールは幾つかのサーバーを表し、それぞれは特定の仕事種類を行わせる為に十分な好み値を有する。サーバーはサーバープールに属していなくても1もしくは複数に属していてもよい。一般的な仕事分配手段の例として、仕事が到来する電話呼びであるような、自動呼びディストリビューター（ACD）がある。

【0022】各サーバーは異なる仕事種類に対して好み値およびしきい値を有する。好み値の大きさは仕事種類に対するアフィニティ（affinity）を表す。サーバーのしきい値は、仕事種類の仕事を行うやる気の無さを表す。サーバーは通常、好み値がしきい値よりも小さいような仕事種類の仕事を受けない。しかし、サーバー割り当て手段が仕事種類に対する複合好み値がその仕事種類に対するしきい値よりも小さくないと判断すると、サーバー割り当て手段はサーバーが仕事種類に対するサーバープールに含まれることを指示される。

【0023】一実施例に従い、サーバーは全ての仕事種類に対して1つのしきい値のみを有する。また、しきい値とサーバー好み値は、他の自動制御システム、および動的サービスおよび利用特性のよって監督オペレーターからのマニュアル的入力をインテグレートするシステムに基づいて動的であることができる。

【0024】サーバー割り当て手段は、多くのユーザ選択可能な関数を用いてユーザ選択可能な入力から複合好み値を計算することができる。複合好み値には、サービスレベル、待ち行列状態、警告状態、所望サービスレベル、応答遅延、呼び放棄率、オペレータ介入のような多くの入力から計算された仕事種類に対する動的好み値を含むことができる。一実施例において、複合好み値関数の入力には、仕事に対するサーバーの好みおよび仕事種類の属性である動的好み値を含むことができる。動的好み値は、サーバー好みを通常参照せずに、他の仕事種類と比較して特定の仕事種類をサーバーに行わせることに対する仕事処理設備を表す。

【0025】サーバー割り当て手段は、仕事種類に対する動的好み値およびサーバーの好み値を入力として受ける複合好み値の多くの関数のいずれかを用いることができる。これら関数は、仕事到着率、目標達成、サーバー割り当て手段をサポートするのに必要な他の関連情報のような他の情報を入力として受けることができる。単純な複合好み値関数として、仕事種類に対するサーバーの好み値および動的好み値を合計するものがある。サーバーが基本的に品質的にふさわしくない（好み値を有しない）仕事種類は通常（少なくとも好ましい実施例において）、動的好み値に関わらず複合好み値を受けない。

【0026】多くの種類の属性を集散的に考慮することにより、企業の仕事運用上如何に組織することを望むかに従い仕事種類を決めることができる。この仕事種類は、必要とされるタスク製品、通信媒体、地理的、行政管轄的、人口統計的、言語、スキル、サービス種類、顧客プロフィール、セールプロセスにおけるステージ、規制クラスのような組み合わせによって決めることができる。以下の例に従い考えてみる。仕事分配動作は、到着する電話呼び、電子メール、ウェブヒット、郵便局からの郵便、他のメールのようなメディアの種類によって好みを区別することを望むことがある。また、見込み顧客、新しい顧客、古い顧客、失った顧客、望ましくない顧客のような顧客関係の種類に基づいて異なるサーブスプロフィールを与えることを望むことがある。また、特定顧客が住む週にて許可されたサーバーを必要とする特定のセールスタスクのような規制制約に基づいて異なる取り扱いを必要とすることがある。ここにおいて、同じタスクが他の週では何れのサーバーでも取り扱えるような場合もある。

【0027】携帯電話会社への問い合わせにおいて、特定の公共的な規制による異なる処理制約を満足させる為に地理的に分類する場合がある。見込みおよび顧客を多くのマーケティングセグメントに従って分類し、マーケットセグメンテーション戦略を用意にすることができる。年齢、収入、教育度合い、仕事、住居の大きさのような統計的なファクターを用い、多くの製品ラインに対するターゲットマーケットにおいて顧客の所屬を特定することができる。仕事分配には住所変更、課金問い合わせ、ローン依頼、勧誘セール、勧誘しないセール、製品ラインにわたるセール、テクニカルサポート、苦情、問題特定、問題解決のようなタスクの種類による分類を必要とすることがある。このような仕事や顧客属性の組み合わせは通常、仕事種類を非常に複雑にする。本発明は、サーバーを小さな統計的プールへと断片化してしまっても非効率的にはせずにこのような複雑さをもサポートすることができる。

【0028】好ましい一実施例に従い、ユーザが選択可能な複合値関数は動的好み値とサーバーの好み値の両方に対して単調に増加すべきである。即ち、もしサーバー好み値が増加あるいは減少すれば、複合好み値は増加あるいは減少する（他の条件は同じと想定した場合）。同様に、もし動的好み値が増加あるいは減少すれば、複合好み値は増加あるいは減少する（他の条件は同じ場合）。勿論、複合好み値関数は、別の入力あるいはより少ない入力を、サーバー割り当て手段や仕事処理システムの必要性に従って用いることができることを当業者は認識できるであろう。また、これら入力から複合好み値関数を作る多くの適切な方法を認識できるであろう。

【0029】複合好み値の選択を動的に操作するオーバーライドする制御規則を用いることもできる。これによ

り、他の所望特性を達成することもできる。他の所望動作特性としては目標を超えるサービスレベルパフォーマンスの等しい割り当て、サーバーの等しい利用、サービスレベル達成における揺らぎのダンピング、サーバーアイドル度における揺らぎのダンピング、モード依存性振る舞いなどが含まれる。オーバーライド制御規則としては、ユーザ選択可能な複合値の方法を制御する別の層を構成することができる。

【0030】複合好み値において動的コンポーネントを計算する方法およびシステムは米国特許出願第?????、"Hierarchical Service Level Remediation for Competing Classes Based Upon Achievement of Service Level Goals"に記載されている。

【0031】サーバー割り当て手段は、目標達成しないもしくは部分的にしか達成しないことに対するサーバープールにおける拡張を行うために仕事の動的好み要求を適用する。同様に、サーバー割り当て手段は、利用可能なサーバーの割合が不釣り合いと見なされる仕事種類サーバープールの短縮(contraction)を行うことができる。また、好みに合う仕事に対する運用規則を行わせることに關して釣り合うサーバーを用いることを可能にし、運用規則に従って公平に仕事を分配することを可能にする。この事により、認識した結果に基づいてサーバーを割り当てるためにマニュアル的に介入することを支援する。上述のようにマニュアル的なディスパッチング(dispatching)は遅すぎることがしばしばであり、再割り当てされたサーバーが十分に早く好ましい仕事に戻らないとより大きな問題となってしまう。本発明は、仕事設備やサーバー組織に対し予期しない制約無しで、仕事種類に対するバックアップサーバーの自動共有利用を動的に可能としない。また、マニュアル的な介入により行われるサーバーの再割り当ての必要性を削減するために、監視員の責任を減らすことができる。

【0032】本発明は、サービスレベル目標を達成すべき仕事種類に対応するサービス要求を拡散することに特に有用である。上述のように、サービスレベル目標には、企業の運用規則に関連したものを用いることができる。これら仕事種類は、サービスクラスと、本明細書で呼ぶサービス要求のクラスを有する。サービスクラスは単に、目標を達成しようとする幾つかのトランザクションを規定するサービスレベル目標の一部である。サービスをよいまたは悪いというように分類する所望のサービスレベルおよび基準は、サービスクラスの一部ではない。例えば、上級見込み客からの電子メール問い合わせの80%を6時間以内に処理する目標を運用上有するものと想定する。「上級見込み客からの電子メール問い合わせ」は、目標に適したサービスクラスである。サーバーが6時間以内に問い合わせすることは、サービスクラスのメンバーシップに影響を与えず、80%の目標値はサービスクラスの定義の一部ではない。

【0033】仕事種類とサービスクラスの間関係はしばしば複雑であり重要なものである。サービスクラスはお互い排他的である必要はなく、また、集合的に包含的である必要はない。サーバースキルに関連しない属性によりサービスクラスを規定することもでき、また、サーバースキルを規定する属性と同じ属性で規定することもできる。また、複数の目標が1つのサービスクラスに属していてもよい。サーバー好みに関して、仕事種類はサービスクラスよりも適切な特徴であることが多いようにすることもできる。

【0034】仕事種類の動的好み値は仕事分配手段においてサーバー割り当て手段への制御入力として構成するベクトルを構成する。この動的好み値は、サーバー割り当てが機能するように判断に強い影響を与える。仕事分配手段は、好み値の大きさに従ってサーバーの仕事種類への分配にバイアスを与える。仕事種類に関連づけられた動的好み値を用いるマシンの動作無しでは、仕事分配ないしACDでは機能することはできるがサーバープールを自動的に拡張ないし縮小する能力は弱くなってしまう。サーバー割り当て手段を用いるマシンを用いると、仕事分配手段は多くのサービスクラスに関連づけられた目標の達成を可能とする動的サーバープールを享受する。

【0035】例えば、仕事分配手段が通常第1サービスクラスのトランザクションに割り当て可能としてマークされた全てのサーバーを使い切ったような状況ではサーバー割り当て手段の存在下において仕事分配手段は、優先度に関連して他の目標と比較して、第1サービスクラスに関連づけられたサービス組織の目標が十分に危機状態となった場合に更なるサーバーを割り当て可能としてマークすることができる。一実施例に従って、サーバー割り当て手段は、可用サーバープールを明示的に拡張し、また、サーバーの仕事アイテムの分配の制御を仕事分配手段に留まらせる。別の実施例に従い、サーバー割り当て手段は、各サーバーに対して仕事アイテムを選択するプロセスにおいて可用サーバープールを暗黙的に拡張する。動的好み値は境界的(marginal)な意味で仕事分配に影響を与えるのみである。例えば、仕事種類に対する高い動的好み値はサーバーが全く品質的にふさわしくない仕事種類のプールにサーバーを配置することは引き継ぎしない。危機状態にある目標が無ければ、サーバー割り当て手段は全てのサービスクラスに対して0の動的好み値を割り当てることができる。この場合、仕事分配手段は動的条件を補わずにサーバーをマッチさせる。

【0036】危機状態の目標を有する仕事種類のサーバープールの拡張は、サーバー割り当て手段およびそのサーバー割り当て方法に対して典型的なモードであるかもしれないが、資源がサービスレベル目標を超えないように制限する必要もある。全ての目標が満足され、一部の

サービスレベルが他の目標よりも目標を超えているような状況において、運用規則は、より優れたサービスをより均等に分配するためにサーバーを再割り当てすることができる。例えば、動的好みは、正で無く負となることができ、これにより、複合好みはサーバーのしきい値よりも低くなる。複合好みやしきい値よりも低く収まれば、サーバーは、サービスレベル目標が満足されないようなときまで、この仕事種類に対するサーバープールから一時的に除去される。従って、全てのサービスレベル目標が満足される場合では、システムは、仕事種類にまたがって整合性を大きくするように、何れの仕事種類に対する資源を拡張し、資源を縮小することにより仕事の間のバランスを取ることができる。

【0037】図1は、階層的仲裁手段102が実行する汎用コンピュータシステム100の高レベルブロック図である。コンピュータシステム100は仕事処理設備にて收容することができる。また、仕事処理設備自体がコンピュータシステム100内に收容されていてもよい。コンピュータシステム100は、中央処理装置(CPU)110、入力/出力デバイス120、コンピュータメモリー130を備える。入力/出力120として、ハードディスクドライブのような記憶装置121、CD-ROMのようなコンピュータが読み取り可能な媒体として提供される階層的仲裁手段102の態様を含むソフトウェア製品をインストールするようにできるコンピュータが読み取り可能なメディアドライブ122がある。入力/出力120はトランザクションを受けるサーバーやトランザクションソースへの接続123を有する。例えば、コンピュータシステム100をコンピュータテレフォニーインタフェース、ウェブサーバ、ウェブブラウザ等に接続するのに接続123を用いることができる。階層的仲裁手段102は、サーバー可用性を検出するため、アプリケーションやトランザクションをサーバーに割り当てるため、サーバーによるトランザクションの処理をモニターするためにこの接続を用いることができる。メモリー130は階層的仲裁手段102を包含することが好ましい。

【0038】上述のように構成したコンピュータシステムにて階層的仲裁手段102を実装することが好ましいが、異なる構成のコンピュータシステムにても実装することができる。例えば、階層的仲裁手段102は、上述のものよりも異なるコンポーネントを有するコンピュータシステム上で実装することができる。また、階層的仲裁手段102はコールセンターにおけるコンピュータシステムのような特殊用途コンピュータシステムにて実装することができる。更に、階層的仲裁手段102は、ASIC等を用いて特殊用途ハードウェアにてソフトウェア無しで実装することができる。コールセンターにおける自動呼び分配(ACD)は、その機能として階層的仲裁手段102を收容することができる。

【0039】階層的仲裁手段102をより詳細に説明するため、以下にて特定の例に従ってその動作を説明する。この例は、階層的仲裁手段102の動作を議論する便利な基礎を提供するように設計したので、階層的仲裁手段102の通常のアプリケーションを必ずしも全て表しているものではない。この例におけるシナリオよりも拡張性があり分散したシナリオにて階層的仲裁手段102を用いることができることを当業者は認識できるであろう。

【0040】図2は、仕事処理設備にて行われる仕事種類を表す多くの仕事種類202に対する多くのサーバーの好みを表すサーバー好み表200である。システムにおける各サーバーはサーバー10001を有する。各サーバーは、サービス組織によって特定される可能性のある仕事種類202それぞれに関連づけられた仕事を理論的には行うことができる。

【0041】各サーバーは、特定の仕事種類202の仕事を行うことに対して好み値203を有する。203の大きさは関連する仕事種類202に対して親和性を表す。また、203の大きさは、202に関連するサーバースキルの尺度を表す。各サーバーは、特定の202に対応するタスクを行うことにするサーバーの抵抗性を表すしきい値204を有する。204は、特定の202に関連するタスクを行うサーバーの能力の無さの尺度をも表す。従って、200は、各サーバー1001、各サーバー仕事種類202に対応するエントリー205~209を有する。例えば、エントリー205は仕事種類「見込み客からの呼び」に関連して100を有するサーバーの好みを表す。この特定の仕事種類に対して、100は「8」のサービスクラスの好み値、「9」のサービスクラスのしきい値を有する。サーバーは通常、好み値がしきい値よりも小さいような仕事は行わない。従って、100は通常、「見込み客からの呼び」に関連づけられたタスクを行わない。他方、エントリー205に示すように、100は通常、「苦情呼び」206に関連づけられたタスクを行う。なぜなら、この仕事種類に対する203は「4」であり204は「3」であるからである。

【0042】サーバーが基本的にふさわしくない仕事種類は、好み値を有しないか「0」の好み値を有する。例えば、サーバー101はエントリー209に示すように「優先顧客からの呼び」に関連づけられた仕事を行うのに品質上ふさわしくない。従って、仕事分配手段は、優先顧客からの呼びに関連づけられた仕事を行わせるように101を割り当てない。

【0043】図3は、複合好み値に関連して仕事処理設備にて行われる仕事種類のサーバー好みおよびしきい値を示す表300を示す。特定仕事種類に対する個々の好みを表すだけでなく、仕事種類に関して仕事処理設備の好みをも表す。図2と図3を比べると、表300に於

ける列301~304の内容がそれぞれ表200に於ける列201~204の内容に対応することが解る。また表300は、複合好み値列305の更なる列を有し、これもまた、仕事種類302を含む仕事をサーバーが行うのに対しての仕事処理システムの全体の好みを表す。例えば、100の「見込み客からの呼び」に対する複合好み値は「10」であり、「苦情呼び」に対する100の複合好み値は「8」である。

【0044】102は、仕事種類に対する動的好みを含む複合好み値を計算するための多くの関数を用いることができる。仕事種類に対して動的好み値を生成する手段の少なくとも1つは、米国特許出願第09/149877 "Dynamically Assigning Priorities for the Allocation of Server Resources to Competing Classes of Work Based Upon Achievement of Service Level Goals," filed on September 8, 1998 に記載されている。

【0045】図4は、102を用いた複合好み値の生成ないしそれに対応する方法を示す流れ図である。

【0046】102は、選択されたアイテムの仕事種類に対するサーバーの好み値を配置する(401)。本発明のある実施例に於いて、102は、選択された仕事アイテムに対する仕事種類を決めるようにすることもできる。また、選択された仕事アイテムに対する仕事種類の決定は、複合好み値の生成の前に行うこともでき、102は仕事アイテムに対する決定した仕事種類を単に配置する。102は、図2に示すサーバー好み表200のようなサーバー好み表を調べることができる。可用サーバーが図2に示したサーバー好み表202のサーバー10100であり選択された仕事種類が「見込み客からの呼び」であると想定する。従って、102は、「見込み客からの呼び」に対する100の好み値が好み値「8」を有するものと特定する。

【0047】102は仕事アイテムに対する動的好み値を配置する(402)。前述のように、動的好み値は仕事種類の属性とすることができる。仕事種類に対する動的好み値を決定するのに多くの関数を用いることができる。例えば、動的好み値が、獲得したサービスレベル、待ち行列状態、所望サービスレベル、警告状態、応答遅延、呼び放棄率、オペレータ介入等に基づくようにすることができる。動的好み値が所望サービスレベルの獲得／非獲得および仕事種類に対する優先度の組み合わせに基づくことと想定する。従って、所望サービスレベルを現在獲得していないより高い優先度の仕事種類は、他の仕事種類(例えば、サービスレベルが獲得しているような仕事種類あるいは低い優先度種類)よりも高い動的好み値を有する。また、「見込み客からの呼び」に対する動的好み値が「4」であると想定する。

【0048】102は、仕事処理設備内で現在選択されている複合好み値関数を配置する(403)。前述のように、102は、複合好み値関数のような多くのパラメ

ーターのユーザ選択を可能にする。多くの場合に、例えば、設備管理において、ある複合好み値関数が他よりも好ましいと判断することができる。

【0049】複合好み値関数において、仕事アイテムの仕事種類に対するサーバーの好み値を仕事アイテムに対する動的好み値と加えることができる(401、402)。また、複合好み値関数は、仕事アイテムの仕事種類に対するサーバーの好み値を仕事アイテムに対する動的好み値と掛けることができる(401、402)。更に、複合好み値関数は、別のパラメーターに従って重み付けされた仕事アイテムの仕事種類に対するサーバーの好み値を、別のパラメーターに従って重み付けされた仕事アイテムに対する動的好み値と加えることができる(401、402)。複合好み値関数において、多くのスケーリングファクターを用いることもできる。

【0050】選択した複合値関数に関わらず、102は複合好み値を計算するために複合好み値関数を用いることができる(404)。選択した複合好み値関数がサーバーの好み値を動的好み値と加えることを伴うものと想定する。「見込み客からの呼び」(8)に対する100の好み値を「見込み客からの呼び」(4)と加えると「12」を得る。

【0051】前述のように、仕事処理設備は通常、サーバーの好み値が同じ仕事種類に対するサーバーのしきい値より小さい場合にはサーバーに仕事を与えない。「見込み客からの呼び」(8)に対する100の好み値は図2に示すように100の対応するしきい値(9)よりも小さい。仕事処理設備は通常、「見込み客からの呼び」に関連する仕事を割り当てるようなサーバーのプール内に100を考慮しない。しかし、上に示したように、100の複合好み値は「12」であり、これは100のしきい値「9」を越える。即ち、100の仕事処理設備が「見込み客からの呼び」ヲ処理する必要性がこのようなタスクを行うことに対する100の抵抗性を現在上まわっている。従って、仕事処理設備は、100がサーバープール内に現在あるものと考慮し、これは、見込み客からの呼びを受ける。勿論、「見込み客からの呼び」仕事種類に対する複合好み値を再計算した場合に100のしきい値よりも複合好み値が低くなれば、100が「見込み客からの呼び」サーバープールから除かれることがあり得る。また、複合好み値は、品質的にふさわしくないサーバーをサーバープールに絶対に割り当てないようにすることもできる。サーバーが基本的に品質的にふさわしくない(好み値が「0」)仕事種類は、複合好み値「0」を受ける。

【0052】複合好み値関数は、サービスクラス好み値および仕事種類に対するサーバーの好み値の多くの関数の何れであってもよい。複合好み値関数は、通常のサービス目標とは独立の多くの動作モードおよび多くの他の動作状態に偶然的ではないように動的に選択することが

できる。このようなモードは、サーバー全体の効率、公平性の考慮、仕事の繁栄要求、または特定の例外的状態に関連する二次的目標に関連していてもよい。動作状態の単純な例として一日に於ける時間があり、別の例として、一週間に於ける曜日がある。ここにおいて、仕事は異なる仕事の規則で運用される。

【0053】図5A、Bは、一実施例に従うサーバーディスパッチ手順を示す流れ図である。サーバー割り当て手段102は検査間隔をセットする(501)。この検査間隔はいずれかの時間間隔または条件を含むようにできる。少なくとも一実施例に於いて、検査間隔は、仕事処理設備内の他の機能によってセットされる。

【0054】102は、終了条件になったかどうか判断する(503)。終了条件は幾つかのファクターによって発生することができ、これには仕事処理設備のシャットダウンを含む。もし終了条件になれば(503)、102は動作を終了する。そうでなければ102は終了条件が発生するまで何らかの規則的な間隔で動作を継続する。

【0055】もし終了条件が発生しなければ(503)、102は検査間隔が終わったかどうかを判断する(505)。もし検査間隔が終わっていないならば(505)、102は再び終了条件が発生したかどうか(503)および/または検査間隔が終わったかどうか(505)を判断するまで待つ(507)。検査間隔の長さはユーザが選択可能にすることができ、検査間隔は実質的に継続であるかのように頻繁にすることもできる。

【0056】もし検査間隔が終われば(505)、102は全ての仕事種類が検査がされたかどうかを判断する(509)。もし全ての仕事種類が検査されていれば(509)、102は終了条件が発生したかどうかを判断するように戻る(503)。済んでなければ、別の検査間隔が終了したかどうかを判断するに進む(505)。

【0057】勿論、102は通常、常に全ての仕事種類と全てのサーバーを解析するのではないことを認識できるであろう。実際には、「検査」はサーバーが可用となったときに開始することができる。検査には、そのサーバーのみを考慮し、すでに待ち行列に入れられた仕事アイテムによって表される仕事種類にのみ関する。また、検査は仕事アイテムが仕事処理設備に到着したときに開始させることができる。検査間隔は、新しく到着した仕事アイテムを考慮し、すでに仕事を待っているサーバーに関連する。好み値が増えたりまたはしきい値が減ることにより、検査をトリガーさせることができる。サーバーは一般に、サーバーが他の割り込むことができない仕事のお陰で現在ビジー状態であるときに新しいサーバープールに入ることを必要としない。勿論、現在待ち行列に何れの仕事も入れられていない場合に仕事種類のサーバープールに於いてサーバーのメンバーシップを判断す

る必要性は少しある。一実施例に従い、これら条件が 102 に関連する処理要求を相当に減らすことができることを認識することができるであろう。

【0058】もし全ての仕事種類が検査されていなければ(509)、102は検査する次の仕事種類を選択する(511)。102は全てのサーバーが検査されたか判断する(513)。もし102が全てのサーバーを検査したならば(513)、102は全ての仕事種類が検査されたかどうかを判断することに戻る(509)。

【0059】もし全てのサーバーが検出されていなければ(513)、102は次のサーバーを選択する(515)。102は、選択されたサーバーの好み値がサーバーのしきい値よりも小さいか判断する(517)。もし選択された好み値がサーバーのしきい値よりも小さくなければ(517)、102はそのサーバーが選択した種類に対して可用とする(523)。選択した仕事種類に対してサーバーを可用とすることには、仕事処理システムにて仕事割り当て設備によって用いられるサーバープールを変更することを伴うことができる。場合によってはサーバーはその仕事種類に対して前にサーバープールに割り当てられていて、更なる処理が必要ない場合もある。勿論、上述のように、全てのサーバーや全ての仕事アイテムが常に検査される必要はない。

【0060】もし選択されたサーバーの好み値が選択された仕事種類に対するサーバーのしきい値よりも小さければ(517)、102は仕事種類とサーバーに対応づけられた複合好み値を更新する(519)。図4は、複合好み値を更新する手順を示している。仕事種類とサーバーに対応する複合好み値を更新することには、102が用いるべき複合好み値関数を決定することができる。上述のように、ユーザーは102により用いられた複合好み値関数を選択することができる。

【0061】102は選択した複合好み関数を適用し、複合好み値がサーバーのしきい値以上であるかを判断する(521)。もし複合好み値がサーバーのしきい値以上であれば、102はその仕事種類に対してサーバーを可用にする(523)。102は単に仕事処理設備に対してサーバーが仕事種類に対するサーバープールに含まれるべきであることを示す。これは、実際にサーバーをその仕事種類に対するサーバープールに割り当てるのに必要なステップを行わなくても行わずに行われる。もし複合好み値がサーバーのしきい値以上でなければ、102はその仕事種類に対してサーバーを選択せず、他のサーバーの検査に戻る(513)。

【0062】サーバーディスパッチ手順のために図5A、Bに示した流れ図は、中央化したサーバーと分散したサーバーの両方のディスパッチ手順に等しく適用可能である。従って、そのサーバーディスパッチ手順は中央化した環境でも分散した環境でも存在することができる。例えば、動的好みの評価によっては、1もしくは複

数のプロセッサで動作する1つの処理によって決定することもでき、また、複合好みはそれぞれがあるサーバーに専用であるプロセッサによって決定することができる。ここで、これら専用プロセッサは何らかの通信手段によって幾つかの動的好みを獲得する。サーバーディスパッチプロセスの実際の処理は、特定の実装に於いて実用的であるいずれの構成によっても実現することができる。

【0063】図6A~Dは、102によって、多くのサーバープール603~605に対するサーバーの割り当てにおける時間にわたっての変化を示している。簡単にするため、サーバープール603に対応する仕事種類601が604に対応する仕事種類よりも高い平均複合好み値を有し、604に対応する仕事種類は、605に対応する仕事種類よりも高い平均複合好み値を有するものと想定することができる。

【0064】仕事処理設備にて必要性が満足されないので102は目標が満足されない仕事種類に対してより多くのサーバーを可用にする。サーバーの割り当てに於ける変化は図6Aに示す図から図6Bに示す図へとサーバープールに於ける可用サーバー602の数を変化させる。図6Aにおいて、仕事種類「見込み客からの呼び」は42の可用サーバーのサーバープール603を有する。「見込み客からの呼び」の目標は達成されていないので、図6Bに示すように102は603に於けるサーバーの数を48サーバーへと増やして更新する。勿論、1つのサーバーには複数のサーバープールを割り当てるようにしてもよい。

【0065】同様に、「苦情呼び」サーバープール604におけるサーバーの数は、図6A、Bで表される時間間隔の間に減少している。「苦情呼び」サーバープール604に於けるサーバーの数の減少は、サーバー好み値の変化、あるいは動的好み値に於ける変化であることがある。「優先顧客からの呼び」仕事種類に対するサーバープール605は、図6A、Bで表される時間間隔の間に一定に留まる。勿論、図6Bで示した「優先顧客からの呼び」サーバープール605におけるサーバーは、図6Aの「優先顧客からの呼び」サーバープール605に対して示したサーバーと同じサーバーである必要はない。

【0066】仕事処理設備に於ける必要性が満足されないので、102は、図6Bと図6Cを比較し示されるように、「見込み客からの呼び」サーバープール603に対してより多くのサーバーを可用とするように動的好み値の利用を継続する。同様に、「苦情呼び」サーバープール604と「優先顧客からの呼び」サーバープール605は集合的に図6B、Cに表した時間間隔の間に減少する。

【0067】102は、サーバープールに対して最大および/または最小の数を保持するように設計することが

できる。例えば、「優先顧客からの呼び」サーバープール605は、最小サーバー数「10」を有することができる。従って、102は、605かその最小サーバー数に到達すると（これは予約サーバーとも呼ばれる）、

「優先顧客からの呼び」サーバープール605からサーバープール603～604へとサーバーを再割り当てすることを控える。

【0068】102は可用サーバーの供給が不釣り合いとみなされる仕事種類のサーバープールを縮小することができる。危機状態の目標を有する仕事種類のサーバープールの拡張が102に対して典型的なモードであるので、102はサーバーレベル目標を超えることからサーバー資源を制限する必要性が出てくる。全ての目標が満足され、一部のサービスレベルは他のサービスレベルが目標を超えるよりも多く越えているような状況において102の運用規則はサーバー資源をより均等に分配するようにサーバーを再割り当てすることができる。

【0069】例えば、図6Cに示す「見込み客からの呼び」サーバープール603がサービス目標を達成するのに必要な量より多くのサーバーを含むと想定する。「見込み客からの呼び」サーバープール603に対応する目標が所望レベルを越えて現在達成されているので、他のサービス目標もまた達成されるようにサーバーの数を減らすことができる。従って、603に於けるサーバーの数は図6Cから図6Dへと減る。サーバープール604～605に割り当てるサーバーの数も図6Cから図6Dへと増える。

【0070】603に於けるサーバーの数を減らす為に102は603に対する動的好み値が正ではなく負となるようにすることができる。これにより、複合値は1もしくは複数のサーバーしきい値よりも低くなる。複合好み値がサーバーのしきい値よりも下に収まれば（サーバーの好み値もまたしきい値よりも下である）、ここでは、サービス目標が満足されなくなるまでサーバーはその仕事種類に対するサーバーから一時的に除去される。従って、全てのサービス目標が満足される場合では、102は、幾つかの仕事種類の組の間で大きな整合性を与えるように何れの1つの仕事種類に対する資源（例えば、サーバー）を拡張し、資源（例えば、サーバー）を縮小することにより仕事種類の間のバランスを取ることができる。動的好み値は0に留まっても増加してもあるいは複合好み値を駆動する関数が複合好み値をサーバーのしきい値よりも下にしている間には考慮しないようにすることもできることを認識できるであろう。

【0071】図7は、動的好み値列701、仕事処理設備にて行われる仕事種類に対するサーバー好みとしきい値複合好み値を含む表700を示している。図3の表300と700を比べると、表300の列301～305が表700の列301～305が対応することが解る。また、表700は多くの仕事種類のパフォーマンスに対

するサービス組織の現在の好み値を表す動的好み値列701を含む。

【0072】上述のように、305に記録される複合好み値を計算するのに多くの方法を用いることができる。表700に於いて、305に於ける複合好み値は303に対するサーバー好みを701に対する動的好み値と合計することによって決定することができる。例えば、100の「見込み客からの呼び」に対する複合好み値は「10」であり、これは、この仕事種類に対する100の好み値「8」をこの仕事種類に対する動的好み値「2」と加えることにより計算される。他の点に関しては上の動作で説明する。

【0073】仕事種類に対するしきい値は特定のサーバーに関係を発生させることと関連して前に説明した。例えば、図2に於いて「苦情呼び」仕事種類に対する100のしきい値は「3」であり、「苦情呼び」仕事種類に対する101のしきい値は「7」である。同様に、「苦情呼び」仕事種類に対する100のしきい値は「見込み客からの呼び」仕事種類に対する100のしきい値とは異なる。

【0074】別の実施例に於いて、仕事種類に対するしきい値は、特定のサーバーを参照せずに確立することができる。図8は、仕事処理設備にて行われる仕事種類を表す多くの仕事種類202に対する多くのサーバーの好みを示したサーバー好み表800である。800は図2に示した200と似ているが800には図2に示した仕事種類列204に対するサーバーしきい値を含まない。

【0075】別の実施例に於いて、800は、多くの仕事種類に対する適切なしきい値を含むしきい値表806と関連づけることができる。この実施例に於いて102は特定のサーバーに仕事を割り当てるかどうかを判断する際にしきい値表806に於いて適切な値を参照する。例えば、しきい値表806は、一実施例に従い、全ての仕事種類に対して1つのみのしきい値を有するようにできる。代わりに、806は各サーバーに対して全ての仕事種類のしきい値を有するようにできる。代わりに、しきい値とサーバー好み値は、他の自動化制御システムおよび動的サービスおよび利用特性を有する監督オペレーターからのマニュアル的な入力をインテグレートするようなシステムに基づいて動的であることができる。また、しきい値は各仕事種類に対して異なるように決めることができる。勿論、800および806は、共通のメモリ素子に記憶することができ、また、相互に関連するデータ構造を有することもできる。

【0076】勿論、サーバー好みの決定に対する仕事種類に貢献する要素は、動的好み値を決定する属性とは全く独立であるようにすることができる。例えば、図8に示すように、100に対してサーバー好みを与えるのに必要な仕事種類が100の位置や発呼者の位置を示す地理的属性のような他の属性と独立であることが容易に認

識できるであろう。しかし、好みの動的成分（これはサーバーとは独立である）は、地理的属性のような他の属性的に決めることができる。また、図8で示す仕事種類は、例えば、好みの動的成分を決めるのにおいて利用する必要は全くない。

【0077】図9Aは別の汎用コンピューターシステム100の高レベルブロック図であり、しきい値判断手段911および複合好み強化手段910を有する。911は、図8に示すしきい値表806の為にしきい値を計算することができる。複合好み強化手段910は現在選択されたユーザ選択可能な複合好み値関数が仕事処理設備に関連して企業の目的を達成するのに最も適した複合好み値関数であるかを判断することができる。他の点では、100は図1に示した100に記載したのと同様に動作する。勿論、複合好み強化手段910と911は102のような他の機能に含まれていてもよい。

【0078】910は、複合値関数の選択を動的に操作して他の所望の動作特性を達成するようなオーバーライドする制御規則を用いることができる。この他の所望の動作特性としては目標を超えたサービスパフォーマンスの均等な割り当て、サーバーの均等な利用、サービスレベル獲得に於けるばらつきの解消、サーバーアイドル具合に於けるばらつきの解消、モード専用の振る舞い等がある。この実施例に於いて、910は、前述の複合好み値関数の上に別の制御層を構成するように考えることができる。

【0079】910は、通常のサービス目標とは独立な多くの動作モードおよび他の動作状態にコンティンジェント（contingent）に複合好み関数を動的に選択することができる。このようなモードは、サーバー全体の効率、公平性の考慮、仕事満足欲求や特定の例外的な条件等を伴う二次的目標に関連していてもよい。

【0080】図9Bは、一実施例に従い、910に行われる幾つかの動作を示す流れ図である。

【0081】910は、企業およびその関連する仕事処理状態を評価し、仕事処理設備に対する企業の目的の全体の達成レベルを判断する（901）。次に910は現在選択されているユーザ選択可能な複合好み値関数が、企業とその関連する仕事処理状態の評価と関連して、適切な複合好み値関数であるかどうかを判断する（902）。

【0082】もし910が現在選択されているユーザ選択可能な複合好み値関数は、適切であると判断すれば（902）、910は待ち状態に入り（903）、その後また企業評価を行う（901）。

【0083】もし910が現在選択されたユーザ選択可能な複合好み値関数は適切ではないと判断すれば（902）、910は現在の企業評価に基づいて別のユーザ可能な複合好み値関数を選択する（904）。そして910は待ち状態に入り（903）また企業評価を行う（9

01）。

【0084】待ち状態（903）は、多くの手段および多くの理由によって終了させることができる。例えば、待ち状態は所定の時間間隔とすることができる。また、複合好み値関数または他のイベントの利用と経験的（contingent）にないし同期的にトリガーさせることができる。前述のように、多くの種類の促成の集まりは、一実施例に従い、仕事種類を決めることができる。仕事種類は、多くの考慮する点、例えば、必要とされるタスク、製品、通信媒体、地理、行政境界、人口統計、言語、スキル、サービス種類、顧客プロファイル、セールスプロセスに於けるステージ、規制クラスのような組み合わせによって決めることができる。

【0085】図10は、一実施例に従い、複合好み値と組み合わせる仕事処理設備にて行われる仕事種類のサーバー好みおよびしきい値を示す表1000である。1000は一般的に図3に示す300と似ているが、仕事種類列302に於ける仕事種類は2以上の属性を有している。例えば、行1001は複数の仕事種類「西部地域における見込み客からの呼び」を含む。表300に於ける対応する行306は1つの属性の仕事種類「見込み客からの呼び」を含む。仕事種類に複数の属性を割り当てると、システム全体の複雑さおよび仕事種類エントリーの総数を増やすこととなりうる。表1000は100に対する属性のみをリストしているが、他のサーバーにも同様な情報を入れることができる。

【0086】表1000を用いて以下の例を考えてみる移動体電話会社への問い合わせが地理的に分類し異なる処理制約を満足し、また、サーバー好みの判断のための仕事種類としてその問い合わせを分類しようとする。仕事処理設備に入る仕事アイテムは第1属性（例えば、「見込み客の呼び」、「苦情呼び」）のみで分類するのではなく、第2属性（例えば、地理的情報）に従っても分類することができる。また、仕事アイテムを多くのマーケット区分に従って分類し、マーケット区分戦略を促進するように分類することができる。年齢、収入、教育、職業、住居の大きさのような統計学的なファクターを用いて多くの製品ラインのターゲットマーケットに対して顧客を特定するのに用いることができる。通常、仕事と顧客属性の組み合わせによって非常に複雑な仕事種類の組み合わせを発生させる。本発明は、サーバー人口を小さな統計学プールへと不適当に断片化するような非効率的な方法を用いずにそのような複雑さをサポートすることができる。

【0087】表1000は、頻繁には実行されないが高い優先度を有する仕事種類に対する必要性を示している。例えば、仕事処理設備は、行1005に示した「公共利益命令呼び」仕事種類のような多くの規制制約の下で動作することがある。この「公共利益命令呼び」仕事種類は、企業にとっては非常に僅かの利益しかない呼

びであるが、公的に命令された最大／最小レベルにてサービスを受けられるようにしなければならない呼びである。従って、表1000は、行1005にて高い複合好み値を示してあり、「公共利益命令呼び」は公共的に命令されたサービスレベルに到達するまで近くないし少し下になっている。102は「公共利益命令呼び」の必要とされたサービスレベルを再び獲得すれば、複合好み値は高い値から落ちる。規制制約に加えて、表1000は企業では他の余り重要でないような高い値を有する目標をも含むことがある。例えば、多くの電話通信サービス労働者組合との合意によって要求されるサービスレベルの達成等である。

【0088】前述のように、サーバーディスパッチ手順は中央化されていても分散化されていてもよい。図11は、分散型サーバー割り当てシステム1101を示している。この1101は、コンピュータシステム100a～100fからなる。コンピュータシステム100a～100fは、サーバー割り当て手段102a～102fを有し、他は図1に示すコンピュータシステム100に類似している。

【0089】サーバー割り当て手段102a～102fは、102に対して説明した方法で動作することができる。サーバー割り当て手段102a～102fは、1つの仕事処理設備内で位置していてもよく、お互い仕事アイテム（ないし個々の仕事アイテム）を共有する複数の仕事処理設備に持って分散していてもよい。

【0090】サーバー割り当て手段102a～102fの1つがサーバー割り当てタスクを達成することができなければ、サーバー割り当て手段はネットワーク1100を用いて他のサーバー割り当て手段と通信する。サーバー割り当て手段102a～102fはネットワーク1100を用いてお互いと状態情報を通信させることもできる。

【0091】幾つかのサーバー割り当て手段102a～102fは、動的好み値判断を行い、また、他のサーバー割り当て手段102a～102fは複合好み値関数を判断するようにできる。この実施例に於いて、サーバー割り当て手段の間では通信が無くてもよく、動的好み値判断を行う特殊なサーバー割り当て手段との通信の他では、お互い独立に動作する。

【0092】本発明には多くの実施例がある。例えば、サーバー割り当て手段が用いる好み値は、全ての資源を割り当てるのに用いることができる。また、隣接した整数ではない好み値を割り当てることができる。また、隣接した整数ではない好みレベルを仕事処理設備によって割り当てることができ、これは数値ではないサービスレベルとすることもできる。例えば、特定の仕事種類に対するよいまたは悪いサービスを満たすことの基準において、トランザクション属性にブール (Boolean) 関数を用いたり、英数字の識別子（例えば、A1>C3）の形

態を用いたりして、正数ないし十進数の利用を避けることもできる。更に、何れの仕事処理設備に於いてもサーバーの好み値のしきい値、動的好み値に対して、値を決めるのに周知な実験的な手順を用いることもできる。

【0093】本明細書に於いて、サーバーがサーバープールに属するものとして説明したが、サーバーは特定のサーバープールのメンバーとして特定できる必要がないことは認識できるであろう。即ち、サーバーは単にある仕事種類ないし他の種類の仕事を受け、サーバープールが例え物理的には存在しなくても仕事種類に対するサーバープールに属するものと考えることができる。

【0094】本発明は、仕事種類とサービスクラスの両方に関連して仕事アイテムを説明できるようなシステムにもすることができる。サービスクラスは、仕事アイテムの特定の特徴を表すことができ、また、仕事種類はサーバーの観点から仕事アイテムを表すことができる。従って、実施例によっては、サービスクラスは、動的好み値に影響を与え究極的には複合好み値に影響を与える。

【0095】上の実施例に於いて、コールセンターに於ける呼びと関連して説明したが、本発明はコールセンターに於いてサーバーによってサーバーによって処理される呼びに制約されず、人間のおよび／または人間でないものであってもよい。本発明は、何れの形態の仕事種類、何れの形態のサービスクラスのサービスにも適用することができる。例えば、本発明は、電子メールメッセージやビデオストリーミングの処理にも適用することができる。

【図面の簡単な説明】

【図1】本発明の一態様に従って階層的仲裁手段102が実行する汎用コンピュータシステム100の高レベルブロック図。

【図2】仕事処理設備にて行われる仕事種類を表す多くの仕事種類に対する多くのサーバーの好みを表すサーバー好みテーブル200の図。

【図3】仕事種類に対する複合好み値に関連して仕事処理設備にて行われる仕事種類のサーバー好みやしきい値を表す表300の図。

【図4】サーバー割り当て手段102を用いる複合好み値の生成を示す流れ図。

【図5】サーバーディスパッチ手順を示す流れ図。

【図6】サーバーは割り当て手段102によるサーバープール603～605に対するサーバーの割り当ての時間にわたった変化を示す図。

【図7】動的好み値列701を含む表700の図。

【図8】仕事処理設備にて行われる仕事種類を表す多くの仕事種類202に対する多くのサーバーの好みを表すサーバー好み表800の図。

【図9】A しきい値判断手段911や複合好み評価手段910を含む汎用コンピュータシステム100の高レベルブロック図。B 複合評価手段910が行う動作の

一部を示す流れ図。

【図10】複合好み値と関連して別の処理設備で行われる仕事種類に対するサーバー好みおよびしきい値を示す表1000。

【図11】分散型サーバー割り当てシステムの1101の図。

【符号の説明】

400 複合好み値の生成

401 選択した仕事アイテムに対するサーバーの好み値を配置

402 選択した仕事アイテムに対する動的好み値を配置

403 複合好み値関数を配置

404 複合好み値を計算するために複合好み値関数を適用

500 サーバーディスパッチ

501 検査間隔をセット

503 終了条件か？

505 検査間隔か？

507 待つ

509 全ての仕事種類を検査したか？

511 次の仕事種類を選択

513 全てのサーバーを検査したか？

515 次のサーバーを選択

517 選択されたサーバーの好み値<サーバーのしきい値

？

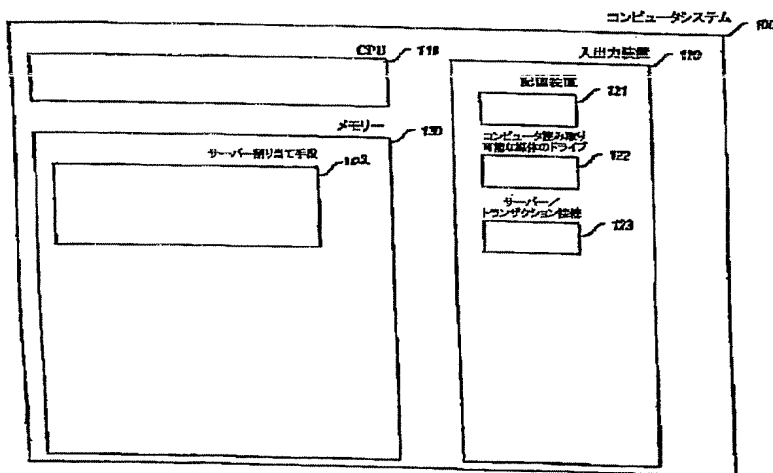
519 複合好みを更新

521 複合好み値 \geq サーバーのしきい値

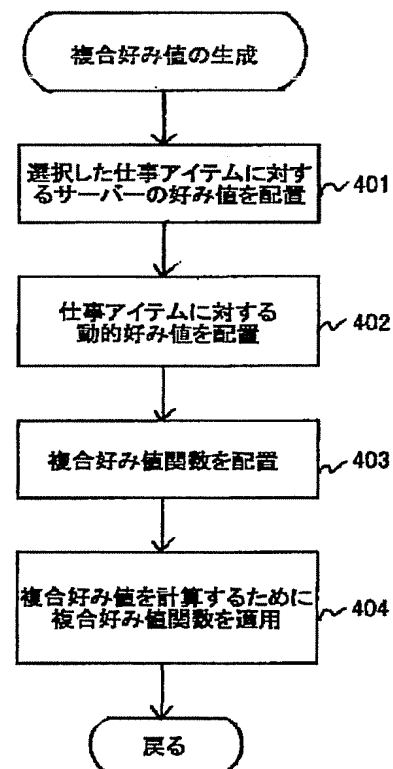
？

523 この仕事種類に対してサーバーを可用にする

【図1】



【図4】



【図2】

201 サーバーID	202 仕事種類	203 仕事種類に対するサーバー好み	204 仕事種類に対するサーバーしきい値	200
100	見込み客からの呼び	8	9	205
100	苦情呼び	4	3	206
101	見込み客からの呼び	5	8	207
101	苦情呼び	6	7	208
101	優先顧客からの呼び	0	0	209

【図3】

301 ID	302 仕事種類	303 仕事種類に対する サーバー好み	304 仕事種類に対する サーバーしきい値	305 仕事種類に対する 複合好み	306
100	見込み客からの呼び	8	8	10	306
100	苦情呼び	4	3	8	307
101	見込み客からの呼び	5	9	7	308
101	苦情呼び	6	7	10	309
101	優先顧客からの呼び	0	0	0	310

【図6】

801 仕事種類	802 可用サーバー	803
見込み客からの呼び	42	803
苦情呼び	30	804
優先顧客からの呼び	15	805

801 仕事種類	802 可用サーバー	803
見込み客からの呼び	48	803
苦情呼び	24	804
優先顧客からの呼び	15	805

801 仕事種類	802 可用サーバー	803
見込み客からの呼び	53	803
苦情呼び	24	804
優先顧客からの呼び	10	805

801 仕事種類	802 可用サーバー	803
見込み客からの呼び	50	803
苦情呼び	28	804
優先顧客からの呼び	11	805

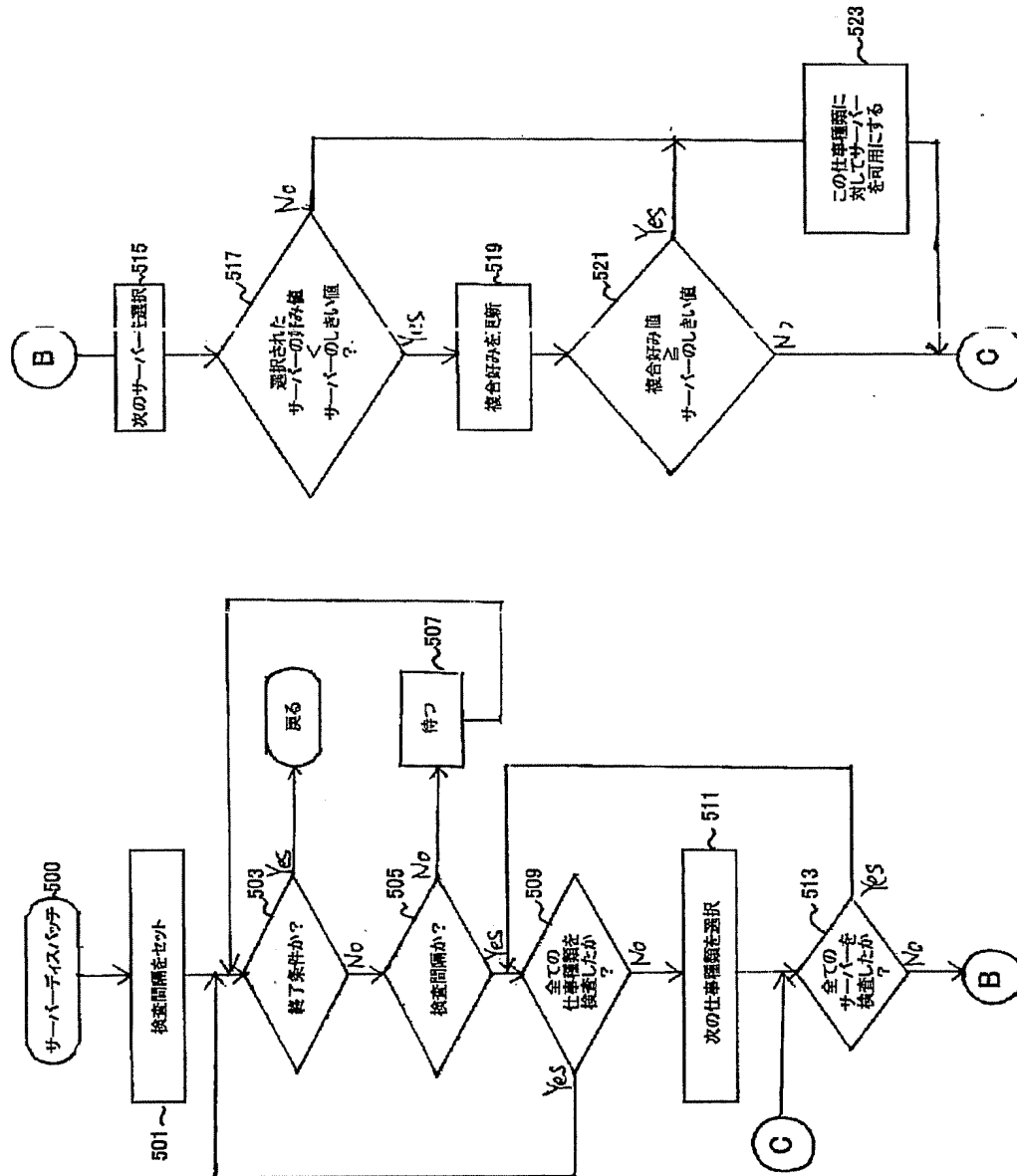
【図8】

201	202 仕事種類	203 仕事種類 に対する サーバー好み	806
100	見込み客からの呼び	8	しきい値値表
100	苦情呼び	4	
101	見込み客からの呼び	5	
101	苦情呼び	8	
101	優先顧客からの呼び	0	

【図10】

301 エージェント ID	302 仕事種類	303 仕事種類 に対する サーバー好み	304 仕事種類 に対する サーバー しきい値	305 仕事種類 に対する 複合好み	1000
100	西部領域の 見込み客からの呼び	8	9	10	1001
100	西部領域から の苦情呼び	4	3	8	1002
100	中央領域の 見込み客からの呼び	5	8	7	1003
100	中央領域から の呼び	6	7	10	1004
100	公共利益命令呼び	8	9	20	1005

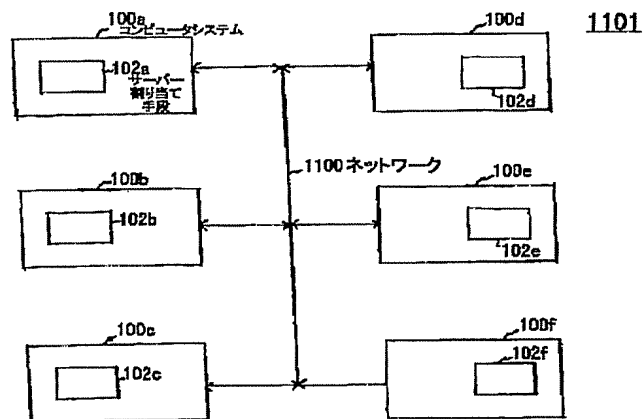
【図5】



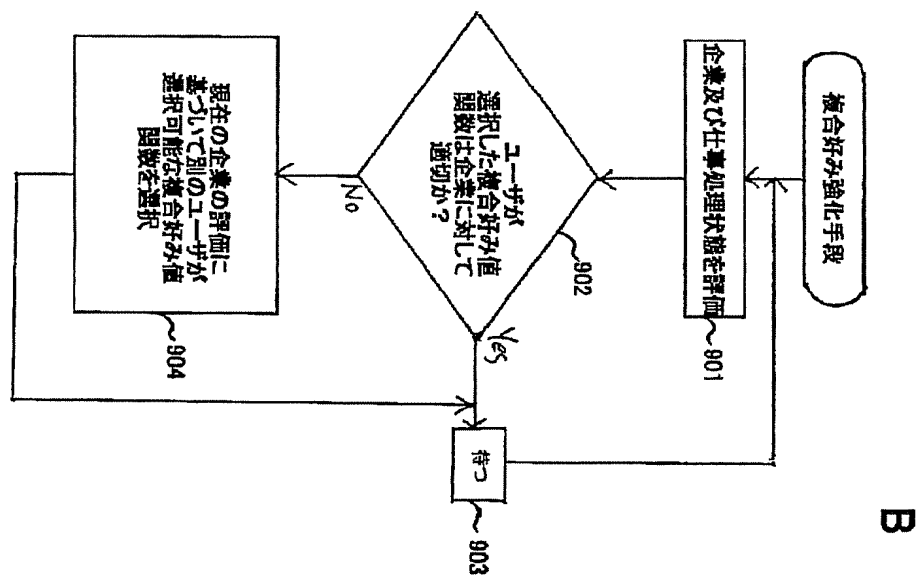
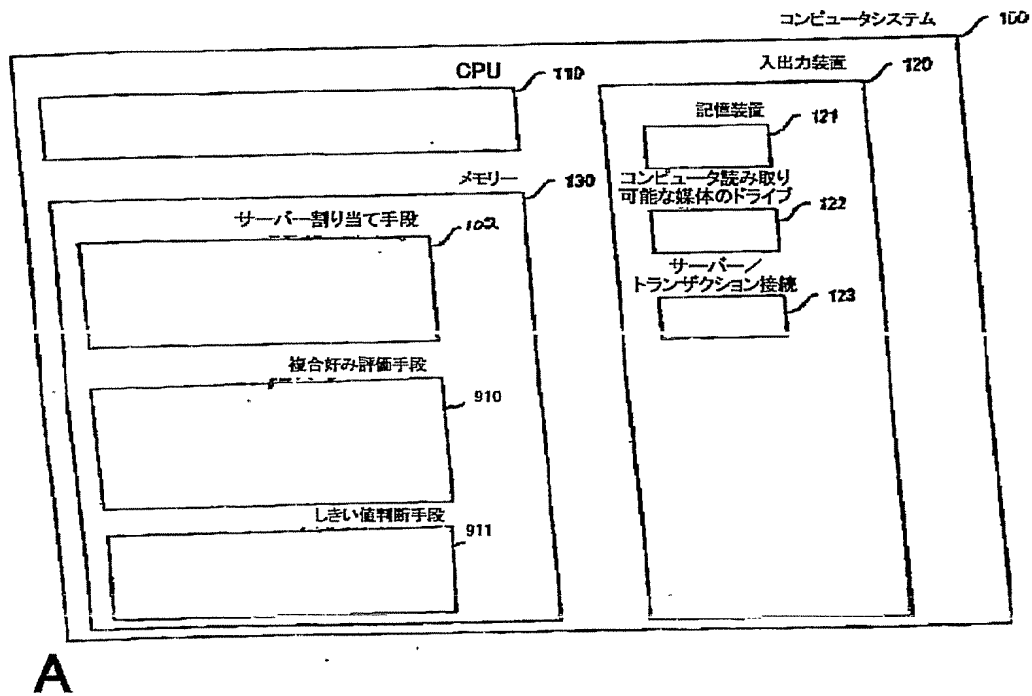
【図7】

301 サーバーID	302 仕事種類	303 仕事種類 に対する サーバー 好み	304 仕事種類 に対する サーバー しきい値	701 仕事種類 に対する 助的好み値	305 仕事種類 に対する 複合好み	700
100	見込み客からの呼び	8	9	2	10	706
100	苦情呼び	4	3	4	8	707
101	見込み客からの呼び	5	9	2	7	708
101	苦情呼び	8	7	4	10	709
101	優先顧客からの呼び	0	0	0	0	710

【図11】



【図9】



フロントページの続き

(71) 出願人 596077259

600 Mountain Avenue,
Murray Hill, New Je
rsey 07974-0636 U. S. A.

(72) 発明者 トーマス ジェー. エドワーズ
アメリカ合衆国、98112 ワシントン、シ
アトル、マドローナ プレース イースト
131

(72) 発明者 デーヴィッド マーレン
アメリカ合衆国、98052 ワシントン、レ
ドモンド、ワンハンドレッドエイティーフ
ァイブス アヴェニュー エヌイー 6464

【外国語明細書】

1. Title of Invention

DYNAMICALLY ALLOCATING SERVER RESOURCES TO COMPETING CLASSES OF WORK BASED UPON ACHIEVEMENT OF SERVICE GOALS

2. Claims

1. A method for determining if a server should be assigned to a server pool for a work type in a work processing facility, the server pool configured to retain an indication of servers that process work of the work type, comprising:

5 determining if a server's first value for the work type is less than a server's second value for the work type, wherein the first value indicates the server's positive indicator for the work type and the second value indicates the server's negative indicator for the work type;

10 determining a composite preference value for the work type if the server's first value for the work type is less than the server's second value for the work type;

determining if the determined composite preference value for the work type is greater than the server's second value for the work type; and

5 sending an indication to the work processing facility that the server is suitable for assignment to the server pool for the work type if the composite preference value for the work type is greater than the server's second value for the work type.

2. The method of claim 1 wherein determining the composite preference value comprises:

determining a dynamic preference value for the work type, wherein the determined dynamic preference value reflects a measurement of the work processing facility's need to have servers perform work of the work type;

accessing a user-selectable composite preference value function that is configured to determine a composite preference value; and

providing the server's determined preference value for the work type and the determined dynamic preference value for the work type to the accessed user.

selectable composite preference value function to generate the composite preference value for the work type.

3. The method of claim 1, further including:

5 not assigning the server to the server pool of the work type if the server's first value is less than the server's second value unless the composite preference value for the work type is greater than the server's second value for the work type.

4. The method of claim 1, further comprising:

0 re-determining the composite preference value for the work type;
determining if the re-determined composite preference value for the work type is greater than the server's second value, and

sending another indication to the work processing facility that the server should be removed from the server pool for the work type if the re-determined composite preference value for the work type is less than the server's second value.

5. The method of claim 1 wherein:

the work processing facility has a plurality of server pools for a plurality of work types, and wherein

a composite preference value is determined for each work type of the plurality of work types if the server's first value for that work type is less than the server's second value for that work type.

6. A system for determining if a server should be assigned to a server pool for a work type in a work processing facility, the server pool configured to retain an indication of servers that process work of the work type, comprising:

a first comparator configured to determine if a server's first value for the work type is less than the server's second value for the work type, wherein the first value indicates the server's positive indicator for the work type and the second value indicates the server's negative indicator for the work type;

a second comparator configured to determine a composite preference value for the work type if the server's first value for the work type is less than the server's second value for the work type;

a third comparator configured to determine if the determined composite preference value for the work type is greater than the server's second value for the work type; and

a result indicator configured to send an indication to the work processing facility that the server is suitable for assignment to the server pool for the work type if the composite preference value for the work type is greater than the server's second value for the work type.

7. The system of claim 6 wherein:

the second comparator comprises

a determiner configured to determine a dynamic preference value for the work type, wherein the dynamic preference value reflects a measurement of the work processing facility's need to have servers perform work of the work type; and

a processor configured to provide the server's determined first value for the work type and the determined dynamic preference value for the work type to a user-selectable composite preference value function configured to generate the composite preference value for the work type.

8. The system of claim 6 wherein:

the result indicator is configured to indicate that the server should not be assigned to the server pool of the work type if the server's preference value is less than the server's threshold value unless the composite preference value for the work type is greater than the server's threshold value for the work type.

〃〃

9. The system of claim 6 wherein:

the second comparator is configured to re-determine the composite preference value for the work type; the system further comprising

a fourth comparator that is configured to determine if the re-determined composite preference value for the work type is greater than the server's threshold value, and

a second result indicator that is configured to send another indication to the work processing facility that the server should be removed from the server pool for the work type if the re-determined composite preference value for the work type is less than the server's threshold value.

10. The system of claim 6 wherein:

the work processing facility has a plurality of server pools for a plurality of work types, and wherein

the second comparator is further configured to determine a composite preference value for each work type of the plurality of work types if the server's preference value for that work type is less than the server's threshold value for that work type.

3. Detailed Description of Invention

TECHNICAL FIELD

The present invention is directed to the field of service provider allocation.

BACKGROUND OF THE INVENTION

Many service organizations need to dynamically allocate their servers in order to attain certain goals. Such allocation is typically performed manually. Servers may include service agents, both human and robotic. Increasingly, external performance measures of service delivered dominate internal cost measures, such as utilization and labor costs. Such external measures often consist of classifying certain transactions into meeting or not meeting desired objectives and determining a proportion of those transactions meeting objectives. Such a proportion is called a service level. The service level is measured over some period of time or over some number of transactions.

Examples of service levels are the percentage of customer problems resolved without further activity, the percentage of dispatched taxicabs that reach the rider within the committed time, the proportion of telephone calls handled by a qualified representative without requiring a transfer or referral to another server, the proportion of telephone calls that can be connected to a server without delay, the proportion of e-mail requests that are answered within 24 hours, the percentage of on-time departures of city buses on a particular bus route on weekdays, the proportion of transactions handled not resulting in a customer complaint, the proportion of preferred customer calls handled by fully qualified servers, the percentage of Spanish customers handled by a server fluent in Spanish, the percentage of telephone calls not abandoned by the customer before connection to a server, the percentage of customer inquiry telephone calls that are not blocked at the

central office switch, the percentage of customer sessions with the self-service World Wide Web pages that are not aborted while waiting for a display, the percentage of customer requests via telephone that can be completed immediately while on the phone, the percentage of loan applications processed within one-half hour from the time of the request, and the percentage of priority telephone calls answered within 8 seconds and handled properly by a qualified server, to name a few.

A service organization's goal for a service level in this context is a particular desired value of the service level. The goal is said to be satisfied if the attained service level is at least as high as the desired service level for the goal. Conversely, the goal is said to be unattained if the realized service level is less than the desired service level. For example, the goal of at least 85% of telephone calls from preferred customers each day being answered within 12 seconds would be attained if, among the telephone calls from preferred customers during the current day, 87% were answered within 12 seconds; inversely, if only 84% of such calls are answered within 12 seconds, the goal would be unattained. In this framework the goal is either attained or not. Moreover, no extra benefit is accrued for attaining a service level much higher than the goal.

A service level goal is one commonly used in criteria for contingency actions, including exceptional allocations of resources. Other relevant performance goals in this context may include maximum values for mean wait times, current queue conditions, minimum number of transactions throughput in a work shift, and measures involving various determinable factors indicative of quality service.

The number of server resources allocated to a type of service often affects the service level achieved for that type of service. When such is the case, the operation can usually reallocate servers to the subject work in order to achieve service level goals. Such reallocation generally incurs opportunity cost; however, since service levels for other work suffers. One can often justify this opportunity cost based on an appropriate priority hierarchy as might be established by the enterprise's operating rules.

For example, suppose servers in a call center can handle both loan servicing and sales servicing transactions. When more servers are assigned to sales activities, sales servicing transactions experience a higher service level or answer delay -- that is, the amount of time required to answer each sales call declines. Meanwhile, the loan servicing calls are not answered as promptly, reducing the service level for loan servicing transactions. The service organization may rationalize this by saying that loan servicing is relatively less important because it is not very likely that an existing customer will switch loan companies, and that the company presently needs to acquire new customers that could easily take their business to a competitor if their calls are not answered promptly. The service organization wants to satisfy the goal of loan servicing, but not at the expense of failing to reach the goal in sales. When the sales goal is not in jeopardy, but the loan servicing is failing to meet its goal, the service organization desires to allocate more resources to loan servicing. The service organization wants to meet both goals, but the sales goal is more important than the loan servicing goal and so may preempt it. That is, if the operation can only meet one goal it should be the sales goal.

Within a set of servicing goals, there may be goals that relate to work having a short "opportunity window" as well as goals for work having a long opportunity window. An example of short opportunity window work is a telephone call, which if not answered in several tens of seconds may be abandoned by the caller with limited patience. An example of long opportunity window work is a letter from a customer regarding a billing adjustment that has until the next billing cycle to complete. Thus, the operation has a very short time frame for meeting service objectives associated with short opportunity window work while the window of opportunity to achieve service objectives associated with long opportunity window work is much broader. Hence there may be productive operational strategies to temporarily allocate more resources who are performing long opportunity window work to short opportunity window work in order to meet the servicing goals for the short opportunity window work.

Many service organizations need to dynamically allocate their servers to achieve desired results. Manual interventions typically effect such reallocations. Often servers are held in abeyance, not available for certain types of work, although they satisfy the skill profile required for the work type. However, if conditions so warrant they might be assigned to the work type. The server is said to be a backup server with respect to that work in such a case where the server is not usually utilized for that type of work.

The desire to allocate more server resources to an activity is typically contingent upon the alternative activities that the server resources can perform and the demand for such alternative activities. Each of these alternative activities is also potentially associated with various service levels, each of which has a goal and a level of attainment. Consequently, the reallocation of resources can depend upon service measures for all alternative work types associated with each of the server resources. Manually performing such a potentially complex allocation function can produce significantly sub-optimal results. Often, manual allocation comes too late and leads to more problems when the reallocated servers are not returned to their preferred work soon enough.

Most automatic call distributors ("ACDs") have a feature that is generically called "call overflow." Call overflow makes a server group available for a call queue contingent upon selected conditions. However, ACDs generally lack facilities for holding a server in abeyance from receiving calls contingent upon appropriate dynamic conditions. Generally, the ACD requires explicit control of the process of getting a call to a server but provides little explicit control of the process of getting a server to a call. However, the server's viewpoint of a work type is often critical in choosing between alternative work in order to maximize the aggregate completion of work.

The advent of skills-based routing, in which the skills of each individual server are considered in allocating servers, complicates the situation. Skills-based routing cannot tolerate simplifying fragmentation of resources into

monolithic pools where distinguishing skills are ignored. For this reason, conventional ACDs as well as workflow automation systems fail to meet this need.

Work distribution systems may force users to manipulate server "skills" in order to effect a reallocation of servers, as discussed above. ACDs and other automatic work distributors may report "service levels" only on skill demands or on some kind of queuing point on a distribution map. However, sometimes a "skill" is actually a type of work instead of an attribute of the server's capabilities. These conventional solutions constitute potentially severe limitations on the monitoring and control of service levels important to the service organization.

In this environment, the service organization wants to provide preferential treatment to work activities in a hierarchy that ensures that the best work item is given to a server in view of service goals and the stated priorities of these goals. An automated system that dynamically expands a pool of servers available for work types based on the attainment or non-attainment of determinable service goals pertaining to the work types in a work processing facility would have significant utility.

SUMMARY OF THE INVENTION

The present invention provides a facility for adjusting a number of servers available for particular work types based on unmet or jeopardized service needs in a work processing facility. Servers may include service agents, both human and robotic. A server assignor and a corresponding server assignment method may each be employed in a work distributor to conditionally adjust the assignment of servers to server pools. An example of a common work distributor where work items are represented by incoming phone calls is an automatic call distributor ("ACD").

The server assignor compares a composite preference value for a work type against each server's threshold value for the work type, according to an embodiment of the invention. When the server assignor determines that the composite preference value is greater than or equal to a server's threshold value,

then the server assignor indicates that the server may be included in the server pool for that work type. Each server has a preference value and a threshold value for different work types. The magnitude of a preference value represents an affinity for the work type. The server's threshold value represents a reluctance to perform work of that work type. Servers do not normally receive work having a work type for which their preference value is less than their threshold value for this work type, unless the server assignor has determined that the server's composite preference value exceeds the server's threshold value for that work type, according to an embodiment of the invention.

The server assignor and corresponding server assignment method may compute the composite preference value from a number of user-selectable inputs and utilize a number of user-selectable functions, according to an embodiment of the invention. The composite preference value may include a dynamic preference value that has been calculated from inputs such as service levels, queue conditions, a desired service level, alarm conditions, answer delays, call abandonment rates, and operator interventions. In one embodiment of the invention, the composite preference value function's inputs comprise the server's preference value for the work type and a dynamic preference value for that work type. Accordingly, the composite preference value in this embodiment considers both the server's specific preferences for various kinds of work and the work processing facility's current general preferences for handling various kinds of work based on broader considerations including goals.

The server assignor may utilize any of myriad composite preference value functions that receive as inputs the work type's dynamic preference value and the server's preference value for the work type, according to an embodiment of the invention. A simple composite preference value function comprises summing the server's preference value for the work type and the work type's dynamic preference value.

The server assignor and its corresponding server assignment method apply a dynamic expediency demand of work to effect an expansion in the server

pool for work types having unmet goals. The server assignor and server assignment method allow a work processing facility to utilize backup servers commensurate with the effect of operational rules for expediting work. The server assignor and server assignment method enable the utilization of backup servers without undue constraints upon the work processing facility and the server organization.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a facility for adjusting a pool of servers available for particular work types based on unmet service needs in a work processing facility in order to maximize the achievement of service goals. Servers may include service agents, both human and robotic. A server assignor and a corresponding server assignment method may each be employed in a work distributor to conditionally adjust the assignment of servers to server pools. A server pool represents a set of servers, each of which has a preference sufficient to render him available to perform a certain work type. A server can belong to none, one, or more server pools. An example of a common work distributor where work is incoming telephone calls is an automatic call distributor ("ACD").

Each server has a preference value and a threshold value for different work types. The magnitude of a preference value represents an affinity for the associated type of work. The server's threshold value represents a reluctance to perform work of that work type. Normally, the server does not receive work for a work type for which the preference value is less than the threshold value. However, when the server assignor determines that the composite preference value for a work type is no longer less than the server's corresponding threshold value for that work type, then the server assignor indicates that the server may be included in the server pool for that work type.

The server may have just one threshold value for all work types, according to one embodiment. Additionally, the threshold values and the server preference values can be dynamically based upon other automated control systems and systems that integrate manual inputs from supervisory operators with dynamic service and utilization characteristics.

The server assignor may compute the composite preference value from a number of user-selectable inputs using a number of user-selectable functions. The composite preference value may include a dynamic preference value for a work type that has been calculated from a variety of inputs, such as service levels, queue conditions, alarm conditions, a desired service level, answer delays, call abandonment rates, and operator interventions. In one embodiment, the inputs of the composite preference value function comprise the server's preference for the work and a dynamic preference value that is an attribute of the work type. The dynamic preference value represents the work processing facility's preference for having servers perform a particular type of work in comparison to other types of work, typically without reference to server preferences.

The server assignor could utilize any of myriad functions for a composite preference value that receive as inputs the dynamic preference value and the server's preference value for the work type. These functions can also receive as inputs other information such as work arrival rates, goal attainment, and other relevant information required to support the server assignor. A simple composite

preference value function comprises summing the server's preference value for the work type and the dynamic preference value. Work types for which the server is basically unqualified (without a preference value) normally (at least in a preferred embodiment) receive no composite preference value regardless of the dynamic preference value.

The confluence of the consideration of many types of attributes can determine the work types depending on how the enterprise's work operation wants to organize. The work type could be determined by a combination of the task required, the product, the communications medium, geographic, governmental jurisdiction, demographic, language, skill, type of service, customer profile, stage in sales process, and regulatory classes. Consider the following examples. A work distribution operation may want to differentiate expediency by type of media such as inbound telephone calls, electronic mail, World Wide Web hits, United States Postal Service mail, and other mail. It may want to provide a different service profile based upon the type of customer relationship such as prospective customer, new customer, old customer, lost customer, and undesirable customer. It could require different handling based upon regulatory constraints, such as particular sales tasks require a server that is licensed in the state where the client resides, while the same task can be handled by any server for customers in other states. An inquiry to a cellular telephone company may be classified geographically in order to satisfy different processing constraints due to regulation by certain public utilities commissions. Prospects and customers could be classified according to various market segments to facilitate a market segmentation strategy. Demographic factors such as age, income, education, occupation, and household size could be used to identify customers in various target markets for various product lines. The work distribution could require classification by the type of task, such as address change, billing inquiry, loan request, solicited sale, unsolicited sale, sale across product lines, technical support, complaint, problem identification, and problem resolution. Typically, combinations of such work and customer attributes effect a very complicated set of work types.

The present invention supports such complications without the inefficiencies of unduly fragmenting the server population into small static pools.

The user-selectable composite value function should be monotonically increasing with respect to both the dynamic preference value and the server's preference value, according to a preferred embodiment of the invention. That is, if the server's preference value increases or decreases then the composite preference value increases or decreases, respectively, with all other things being equal. Likewise, if the dynamic preference value increases or decreases then the composite preference value increases or decreases, respectively, with all other things being equal. Of course, the skilled artisan will recognize that the composite preference value function could utilize additional inputs or even fewer inputs, depending on the needs of the server assignor and the work processing system, and should also recognize a variety of suitable methods for producing a composite preference value function from these inputs. Some embodiments may utilize overriding control laws that dynamically manipulate the selection of the composite value function so as to achieve other desired operating characteristics. Examples of other desired operating characteristics include equal allocation of service level performance in excess of the associated goals, equal utilization of servers, damping of fluctuations in service level attainment, damping of fluctuations in server idleness, and mode-specific behavior. The overriding control laws may constitute another layer of control on top of the method of the user-selectable composite preference value in these embodiments of the invention.

An exemplary method and system for computing a dynamic component in a composite preference value is disclosed in U.S. Patent Application No. 09/248067 "Hierarchical Service Level Remediation for Competing Classes Based Upon Achievement of Service Level Goals," filed on February __, 1999, assigned to the Mosaix Corporation, and which is incorporated herein by reference.

The server assignor and its corresponding server assignment method apply a dynamic expediency demand of work to effect an expansion in the server pool for work types having unmet or marginally achieved goals. Likewise, the server

assignor can effect a contraction of a server pool for work types deemed to have a disproportionate supply of available servers. The server assignor and server assignment method allow a servicing operation to utilize servers commensurate with the effect of operational rules for expediting work and distributing work fairly according to operating rules. This can help eliminate manual intervention to reallocate servers based on realized results. As previously discussed, manual dispatching is often too late and leads to more problems when the reallocated servers are not returned to their preferred work soon enough to meet goals. The server assignor and server assignment method dynamically enable the automatic shared utilization of backup servers for a work type without undue constraints upon the work processing facility and the server organization. The server assignor and server assignment method can improve customer service by meeting service goals and reduce the burden on supervision by reducing the need for server reallocations effected by manual intervention.

The server assignor is particularly helpful in mitigating service demands associated with work types that are subject to service level goals. As previously discussed, the service level goals may relate to an enterprise's operating rules. These work types have classes of service demands that are herein called service classes. A service class is merely the part of the service level goal that defines the set of transactions that apply to the goal. The desired service level and the criteria for classifying the service into good or bad are not part of the service class. For example, assume an operation has the goal of processing 80% of the e-mail inquiries from premium prospects within 6 hours. The "e-mail inquiries from premium prospects" is the service class that pertains to the goal. Whether or not a server processes an inquiry within 6 hours does not affect membership in the service class, and the 80% goal value is not part of the definition of the service class.

The relationship between work types and service classes is frequently complicated and is certainly not trivial. Service classes are not necessarily mutually exclusive, nor are they necessarily collectively inclusive. Service classes may be defined by attributes not related to server skills or they may be defined by the same

attributes that define server skills. Additionally, more than one goal may pertain to a service class. With regard to server preferences, the work type is frequently a more pertinent characteristic than service class, according to an embodiment of the invention.

The dynamic preference values of the work types constitute a vector that functions as a control input to a server assignor in a work distributor. The dynamic preference values may sway decisions on server assignments to work. The work distributor biases server distribution to work types in accordance with the magnitude of their preference values. Without the operation of a machine that utilizes dynamic preference values associated with work types, the work distributor or ACD can still function, but with a diminished ability for expanding or contracting a server pool automatically. With a machine using the server assignor, the work distributor enjoys a dynamic server pool that enables the attainment of the goals associated with various service classes.

For example, in a situation in which a work distributor would normally have exhausted all servers marked as assignable to transactions of a first service class, in the presence of the server assignor, the work distributor may have additional servers marked as assignable once the service organization's goals associated with the first service class become sufficiently jeopardized, relative to other goals in terms of priority. According to an embodiment of the invention, the server assignor explicitly expands the pool of available servers, while the control of distributing work items to servers remains with the work distributor. According to another embodiment of the invention, the server assignor implicitly expands the pool of available servers in the process of selecting work items for each server. The dynamic preference values only affect work distribution in a marginal sense. For example, a high dynamic preference value for a work type will still not place a server in the pool for a work type for which he is totally unqualified. When no goals are in jeopardy, the server assignor may assign a dynamic preference value of zero for all service classes. In this case, the work distributor matches servers without any compensation for dynamic conditions.

While expansion of the server pool for work types having jeopardized goals may be a typical mode for the server assignor and its server assignment method, there may also be a need to limit resources from exceeding a service level goal. In the situation where all goals are being met, and some service levels are exceeding their goals much more than other goals, then the operating rules may reallocate the servers in order to more evenly distribute the superior service, according to an embodiment of the invention. For example, the dynamic preference may become negative, rather than positive, causing the composite preference to be lower than a server's threshold. When the composite preference falls below the threshold, then the server will temporarily be removed from the server pool for this work type until such a time that the service level goal becomes unmet. Thus, in the case where all service goals are being met, the system can balance between work by expanding resources and contracting resources for any one work type to provide greater consistency across work types.

Figure 1 is a high-level block diagram of a sample general-purpose computer system 100 upon which a server assignor 102 may execute, according to an embodiment of the invention. The computer system 100 may be contained within a work processing facility, according to an embodiment of the invention. In another embodiment of the invention, the work processing facility may itself be contained within the computer system 100. The computer system 100 contains a central processing unit (CPU) 110, input/output devices 120, and a computer memory (memory) 130. Among the input/output devices is a storage device 121, such as a hard disk drive, and a computer-readable media drive 122, which can be used to install software products, including server assignor 102, which are provided on a computer-readable medium, such as a CD-ROM. The input/output devices 120 further include a connection 123 to servers and transaction sources from which transactions are received. For example, the connection 123 may be used to connect the computer system to one or more computer telephony interfaces, World Wide Web servers, or World Wide Web browsers. This connection is usable by the server assignor 102 to examine servers, make servers available for applications and

transactions, and monitor the processing of transactions by servers. The memory 130 preferably contains the server assignor 102.

While the server assignor 102 is preferably implemented on a computer system configured as described above, those skilled in the art will recognize that it may also be implemented on computer systems having different configurations. For example, the server assignor 102 may be implemented on a computer system having different components than described above. The server assignor 102 may also be implemented on special-purpose computing systems, such as those in a call center. The server assignor 102 may further be implemented without software in special-purpose hardware, using, for example, application-specific integrated circuits ("ASICs"). An automatic call distributor ("ACD") in a call center or a work flow automation system may also include the server assignor 102 as part of its functionality.

In order to more fully describe the details of the server assignor 102, its operation in conjunction with a specific example is discussed hereafter. This example is designed to provide a convenient basis for discussing the operation of the server assignor 102, and therefore is not necessarily representative in all senses of a typical application of the server assignor 102. Indeed, those skilled in the art will recognize that the server assignor 102 may be applied to scenarios that are both more extensive and diverse than the scenario portrayed in the example.

Figure 2 is a server preference table 200 illustrating the preferences of various servers for the various work types 202 that represent types of work performed in the work processing facility, according to an embodiment of the invention. Each server in the system has a server ID 201. Each server may theoretically perform work associated with each possible work type 202 identified by the service organization.

Each server has a preference value 203 for performing work of a particular work type 202. The magnitude of a preference value 203 represents an affinity for the associated work type 202. The magnitude of the preference value 203 may also represent a measure of a server's skills with regard to the associated work

type 202. Each server has a threshold value 204 that represents the server's reluctance for performing tasks associated with a particular work type 202. The threshold value 204 may also represent a measure of the server's disabilities with regard to performing the tasks associated with a particular work type 202. Accordingly, the server preference table 200 has entries 205-209 corresponding to each server ID 201 and each work type 202. For example, the entry 205 represents the preferences of the server with server ID 100 with regard to the work type "calls from prospects." For this particular work type, the server ID 100 has a preference value for the service class of "8" and a threshold value for the service class of "9." Normally, the server does not do work for which the preference value is less than the threshold value. Accordingly, the server ID 100 would not normally perform tasks associated with "calls from prospects." On the other hand, as indicated in the entry 205, the server ID 100 would typically perform tasks associated with "claims calls" 206 since the preference value 203 for this work type is "4" while the threshold value 204 is "3."

Work types for which the server is basically unqualified have no preference value or a preference value of "0." For example, the server ID 101 is not qualified to perform work associated with "calls from preferred customers" as shown in entry 209. Accordingly, the work distributor will not assign the server ID 101 to perform work associated with calls from preferred customers.

Figure 3 is a diagram showing a table 300 illustrating server preferences and thresholds for work types performed in the work processing facility coupled with a composite preference value, according to an embodiment of the invention. The composite preference value accounts for not only an individual's preferences for a particular work type but also the work processing facility's preferences with regard to the work type. Comparing Figures 2 and 3, it can be seen that the contents of columns 301-304 in the table 300 correspond respectively to the contents of columns 201-204 in the table 200. The table 300 also includes an additional column: a composite preference value column 305 that also represents the work processing system's overall preference for the respective server to perform

work that contains the work type 302. For example, the composite preference value for server ID 100's "calls from prospects" is "10" while the Server ID 100's composite preference value for "claims calls" is "8."

The server assignor 102 may utilize a variety of functions to compute the composite preference value that may include a dynamic preference for the work type. At least one exemplary procedure for generating a dynamic preference value for a work type is disclosed in U.S. Patent Application No. 09/149,877, "Dynamically Assigning Priorities for the Allocation of Server Resources to Competing Classes of Work Based Upon Achievement of Service Level Goals," filed on September 8, 1998, assigned to the Mosaix Corporation, and which is incorporated herein by reference.

Figure 4 is a flowchart illustrating the generation of composite preference values using the server assignor 102 or its corresponding method, according to an embodiment of the invention.

The server assignor 102 locates a server's preference value for a selected work item's work type (step 401). In some embodiments of the invention, the server assignor 102 determines the work type for the selected work item. In other embodiments of the invention, the determination of the work type for the selected work item may be performed prior to the generation of composite preference values, with the server assignor 102 simply locating the determined work type for the work item. The server assignor 102 may examine a server preference table, such as the server preference table 200 shown in Figure 2. Assume that the available server is the server ID 100 shown in the server preference table 200 of Figure 2 and that the selected work type is "calls from prospects." Accordingly, the server assignor 102 would identify the server ID 100's preference for "calls from prospects" as having a preference value of "8."

The server assignor 102 locates the dynamic preference value for the work item (step 402). As previously mentioned, the dynamic preference value could be an attribute of the work type. A number of functions may be utilized in determining the dynamic preference value for the work type. For example, the

dynamic preference value may be based upon attained service levels, queue conditions, a desired service level, alarm conditions, answer delays, call abandonment rates, or operator interventions. Assume that the dynamic preference value is based upon a combination of attainment/non-attainment of desired service level and a priority for the work type. Accordingly, higher priority work types whose desired service levels are not presently being attained will have higher dynamic preference values than other work types, e.g., work types whose service levels are being attained or work types of lower priority. Assume further that the dynamic preference value for "calls from prospects" is "4."

The server assignor 102 locates the composite preference value function presently selected within the work processing facility (step 403). As previously discussed, the server assignor 102 allows for user selection of various parameters, such as the composite preference value function. At various times, facility management, for example, may determine that one composite preference value function is preferable to another.

An exemplary composite preference value function might comprise adding the server's preference value for a work item's work type (step 401) with the dynamic preference value for the work item (step 402). Another exemplary composite preference value function might comprise multiplying the server's preference value for a work item's work type (step 401) with the dynamic preference value for the work item (step 402). Yet another exemplary composite preference value function might comprise adding the server's preference value for a work item's work type (step 401) weighted according to another parameter with the dynamic preference value for the work item (step 402) also weighted according to another parameter. A variety of scaling factors may also be used in still other exemplary embodiments of the composite preference value function.

Regardless of the composite value function selected, the server assignor 102 utilizes the composite preference value function to compute the composite preference value (step 404). Assume that the presently selected composite preference value function calls for adding the server's preference value

with the dynamic preference value. Adding the server ID 100's preference value for "calls from prospects" ("8") with the dynamic preference value for "calls from prospects" ("4") yields "12."

As previously discussed, the work processing facility does not normally provide work to a server when the server's preference value is less than the server threshold value for the same work type. The server ID 100's preference value for "calls from prospects" ("8") is less than the server ID 100's corresponding threshold value ("9"), as shown in Figure 2. The work processing facility would not normally consider the server ID 100 to be in the pool of servers that may be assigned work associated with "calls from prospects." However, as shown above, the server ID 100's composite preference value is 12, which exceeds the server ID 100's threshold value of 9. In other words, the needs of the work processing facility for the server ID 100 to process "calls from prospects" now outweighs the server ID 100's reluctance to perform such tasks. Accordingly, the work processing facility may now consider the server ID 100 to be within the pool of servers who may receive calls from prospects. Of course, when a re-calculation of the composite preference value for the "calls from prospects" work type results in a composite preference value lower than the server ID 100's threshold value, then the server ID 100 may be removed from the "calls from prospects" server pool. In addition, the composite preference value (according to a preferred embodiment) will never result in the assignment of an unqualified server to a server pool. Work types for which a server is basically unqualified (having a preference value of "0") may receive a composite preference value of "0."

The composite preference value function may be any of myriad functions of the service class preference value and the server's preference value for the work type, according to an embodiment of the invention. The composite preference function may be dynamically selected contingent upon various operating modes and various other operating conditions independent of normal service goals. Such modes might be related to secondary goals involving efficiency of the server population, fairness considerations, job enrichment desires, or particular exceptional

conditions. A simple example of an operating condition is time of day while another is a day of the week where the business is operating under different business rules.

Figures 5A and 5B are a flowchart illustrating the server dispatch procedure, according to an embodiment of the invention. The server assignor 102 sets an examination interval (step 501). The examination interval may comprise any number of time periods or conditions. In at least one embodiment, the examination interval may be set by other functionality within the work processing facility.

The server assignor 102 determines whether a termination condition has arisen (step 503). A termination condition may arise due to a number of factors, including the shutdown of the work processing facility. If a termination condition has arisen (step 503), then the server assignor 102 terminates its operations. The server assignor 102 otherwise continues operations at some regular interval until a termination condition arises.

If a termination condition has not arisen (step 503), then the server assignor 102 determines whether an examination interval has concluded (step 505). If an examination interval has not concluded (step 505), then the server assignor 102 waits (step 507) before re-determining whether a termination condition has arisen (step 503) and/or the examination interval has concluded (step 505). The length of the examination interval may be user selectable, and the examination interval may be so frequent as to be essentially continuous, according to at least one embodiment of the invention.

If an examination interval has concluded (step 505), then the server assignor 102 determines whether all work types have been examined (step 509). If all work types have been examined (step 509), then the server assignor 102 returns to determining whether a termination condition has arisen (step 503) before otherwise proceeding to determine if another examination interval has concluded (step 505).

Of course, a skilled artisan will recognize that the server assignor 102 does not typically analyze all work types and all servers all of the time. In practice, an "examination" may begin when a server becomes available. The examination

would concern that server only and with regard only to the work types represented by work items already queued. In addition, an examination may also begin when a work item arrives in the work processing facility. The examination interval would concern the newly arrived work item and with regard to servers already awaiting work. An increase in a preference value or a decrease in a threshold value could also trigger an examination. A server need not generally be included in a new server pool when the server is presently busy with another non-interruptible work item. Of course, little need arises to determine server membership in a server pool for work types where no work is presently queued. As a skilled artisan will recognize, these conditions will substantially reduce the processing demands associated with the server assignor 102, according to an embodiment of the invention.

If all work types have not been examined (step 509), then the server assignor 102 selects the next work type to be examined (step 511). The server assignor 102 determines whether all servers have been examined (step 513). If the server assignor 102 has examined all servers (step 513), then the server assignor 102 returns to determining whether all work types have been examined (step 509).

If all servers have not been examined (step 513), then the server assignor 102 selects a next server (step 515). The server assignor 102 determines whether the selected server's preference value is less than the server's threshold value (step 517). If the selected server's preference value is not less than the server's threshold value (step 517), then the server assignor 102 makes the server available for the selected work type (step 523). Making the server available for the selected work type may entail modifying a server pool utilized by the work assignment facility in the work processing system. On some occasions, the server may have previously been assigned to the server pool for the work type, and no additional processing will be required. Of course, as discussed above, not all servers and all work items necessarily need to be examined all the time.

If the selected server's preference value is less than the server's threshold value for the selected work type (step 517), then the server assignor 102 updates the composite preference value associated with the work type and server

(step 519). Figure 4 provides an exemplary procedure for updating a composite preference value. Updating the composite preference value associated with the work type and server may include determining which composite preference value function the server assignor 102 should utilize. As previously discussed, the user may select the composite preference value function used by the server assignor 102, according to an embodiment of the invention.

The server assignor 102 then applies the selected composite preference value function to determine whether the composite preference value is greater than or equal to the server's threshold value (step 521). If the composite preference value is greater than or equal to the server's threshold value, then the server assignor 102 makes the server available for this work type (step 523). The server assignor 102 may simply indicate to the work processing facility that the server should be included in the server pool for the work type, without actually undertaking any of the steps required to accomplish the assignment of the server to the server pool for the work type. If the composite preference value is not greater than or equal to the server's threshold value, then the server assignor 102 does not select the server for this work type and returns to examining other servers (step 513).

The flowchart shown in Figures 5A and 5B for the server dispatch procedure is equally applicable for both centralized and distributed server dispatch procedures. Accordingly, the server dispatch procedure may exist in a centralized embodiment as well as a distributed embodiment. For example, the evaluation of the dynamic preference can be determined by one process running on one or more processors while the composite preferences could be determined by processors each of which is dedicated to a server, where these dedicated processors obtain the set of dynamic preferences through some means of communication. The actual processing of the server dispatch process may be realized in any configuration that is practical for a particular implementation.

Figures 6A-6D illustrate the change in assignment of servers to various server pools 603-605 over time due to the server assignor 102, according to an embodiment of the invention. For purposes of illustration, one can assume that a

work type 601 associated with the server pool 603 has a higher average composite preference value than the work type associated with the server pool 604 which in turn has a higher average composite preference value than the work type associated with the server pool 605.

Because of unmet needs in the work processing facility, the server assignor 102 allows more servers to be made available for work types whose goals are not being met. Changes in the assignment of servers alters a number of available servers 602 in the server pools from the chart shown in Figure 6A to the chart shown in Figure 6B. In Figure 6A, the work type "calls from prospects" has a server pool 603 of 42 available servers. Because of unachieved goals associated with "calls from prospects," the server assignor 102 increasingly updates the number of servers in the server pool 603 to 48 servers, as shown in Figure 6B. Of course, a server may be assigned to more than one server pool.

Similarly, the number of servers in the "claims calls" server pool 604 has decreased during the time interval represented by Figures 6A and 6B. The decrease in the number of servers in the "claims calls" server pool 604 could be due to changing server preference values or to changes in the dynamic preference value. The server pool 605 for the "calls from preferred customers" work type remains constant during the time interval represented by Figures 6A and 6B. Of course, the servers in the "calls from preferred customers" server pool 605 shown in Figure 6B need not necessarily be the same servers shown for the "calls from preferred customers" server pool 605 of Figure 6A.

Because of unmet needs in the work processing facility, the server assignor 102 continues to utilize a dynamic preference value that results in making more servers available for the "calls from prospects" server pool 603, as shown by comparing Figure 6B with Figure 6C. Similarly, the "claims calls" server pool 604 and the "calls from preferred customers" server pool 605 collectively decrease during the time interval represented by Figures 6B and 6C.

The server assignor 102 may be designed to retain a maximum and/or minimum number of servers for a server pool, according to an embodiment of the

invention. For example, the "calls from preferred customers" server pool 605 may have a minimum server number of "10." Accordingly, the server assignor 102 refrains from re-assigning servers from the "calls from preferred customers" server pool 605 to the server pools 603-604 once the server pool 605 has reached its minimum number of servers, which may also be known as reserved servers.

The server assignor 102 may contract a server pool for work types deemed to have a disproportionate supply of available servers. While expansion of the server pool for work types having jeopardized goals is the typical mode for the server assignor 102 and its server assignment method, the server assignor 102 may also need to limit server resources from exceeding a service level goal, according to an embodiment of the invention. In the situation where all goals are being met, and some service levels are exceeding their goals much more than other service levels are exceeding their respective goals, then the server assignor's operating rules may reallocate the servers in order to more evenly distribute the server resources.

For example, assume that the "calls from prospects" server pool 603 shown in Figure 6C contains more servers than are necessary to achieve the associated service goals. Since the goals associated with the "calls from prospects" server pool 603 are now being achieved beyond a desired level, then the number of servers may be decreased so that other service goals may also be achieved. Accordingly, the number of servers in the "calls from prospects" server pool 603 decreases from Figure 6C to Figure 6D. The number of servers assigned to the server pools 604-605 correspondingly increases from Figure 6C to Figure 6D.

To reduce the number of servers in the server pool 603, the server assignor 102 may make the dynamic preference value for the server pool 603 become negative, rather than positive, causing the composite preference value to be lower than one or more server's threshold value. When the composite preference value falls below a server's threshold value (where the server's preference value is also lower than the threshold value), then the server will temporarily be removed from the server pool for the work type until the service level goal becomes unmet. Thus, in the case where all service goals are being met, the server assignor 102 may balance

between work types by expanding resources (e.g., servers) and contracting resources (e.g., servers) for any one work type to provide greater consistency across some set of work types. It should be obvious to one skilled in the art that the dynamic preference value may remain at zero, be increasing, or not be considered while the function(s) driving the composite preference value cause(s) the composite preference value to fall below a server's threshold value.

Figure 7 is a diagram showing a table 700 that includes a dynamic preference value column 701 as well as server preferences and thresholds for work types performed in the work processing facility and a composite preference value, according to an alternative embodiment of the invention. Comparing the table 300 of Figure 3 with the table 700, it can be seen that the columns 301-305 of table 300 correspond respectively to the columns 301-305 of the table 700. The table 700 also includes the dynamic preference value column 701 that represents the service organization's current preference value for performance of various work types.

As previously discussed, a number of methods may be used to compute the composite preference value recorded in the composite preference value column 305. In the table 700, the composite preference values in the composite preference value column 305 have been determined by adding together the server preference for the work type column 303 with the dynamic preference value for the work type 701. For example, the composite preference value for server ID 100's "calls from prospects" is "10" which has been computed by adding the Server ID 100's preference value of "8" for this work type with the dynamic preference value for this work type of "2." In all other respects, the invention operates in the manner described above.

The threshold value for a work type has previously been described as bearing a relationship to a particular server. For example, as shown in Figure 2, the Server ID 100's threshold value for the "claims calls" work type is "3" while the Server ID 101's threshold value for the "claims calls" work type is "7." Similarly, the Server ID 100's threshold value for the "claims calls" work type differs from the Server ID 100's threshold value for the "calls from prospects" work type.

In an alternative embodiment of the invention, the threshold value for a work type may be established without reference to a particular server. Figure 9 is a server preference table 800 illustrating the preferences of various servers for the various work types 202 that represent types of work performed in the work processing facility, according to an embodiment of the invention. The server preference table 800 resembles the server preference table 200 shown in Figure 2, except the server preference table 800 does not contain the server threshold for work type column 204 shown in Figure 2.

The server preference table 800 may be associated with a threshold value table 806 that contains appropriate threshold values for various work types, according to an alternative embodiment of the invention. In this embodiment, the server assignor 102 will reference an appropriate value in the threshold value table 806 in determining whether to assign work to a particular server. For example, the threshold value table 806 may have just one threshold value for all work types, according to one embodiment. The threshold value table 806 may instead have one threshold value for all work types for each server. Alternatively, the threshold values and the server preference values may be dynamically based upon other automated control systems and systems that integrate manual inputs from supervisory operators with dynamic service and utilization characteristics. In addition, the threshold value may be differently determined for each work type. Of course, the server preference table 800 and the threshold value table 806 may be stored in a common memory element and may have inter-related data structures, according to an embodiment of the invention.

Of course, the elements contributing to a work type for the determination of server preference may be totally independent of the attributes that determine the dynamic preference value, according to an embodiment of the invention. For example, as shown in Figure 8, one can readily see that the work type needed to get the server preferences for the server ID 100 is independent of other attributes, such as a regional attribute describing the location of a caller or the location of the server ID 100. However, the dynamic component of the preference

(which is independent of the server) may be determined partly by other attributes, such as a regional attributes. Additionally, the work types shown in Figure 8, for example, need not necessarily be utilized to any extent in determining the dynamic component of the preference.

5 Figure 9A is a high-level block diagram of another general-purpose computer system 100 that includes a threshold value determiner 911 and a composite preference evaluator 910, according to an alternative embodiment of the invention. The threshold value determiner 911 may compute the threshold values for the threshold value table 806 shown in Figure 8. The composite preference evaluator
1 910 may determine whether the presently selected user-selectable composite preference value function is the most appropriate composite preference value function for achieving the enterprise's objectives with regard to the work processing facility, according to an embodiment of the invention. In all other respects, the computer system 100 operates in the manner previously described for the computer system 100 shown in Figure 1. Of course, the composite preference evaluator 910 and the threshold value determiner may be included in other functionality, such as the server assignor 102.

The composite preference evaluator 910 may utilize overriding control laws that dynamically manipulate the selection of the composite value function so as to achieve other desired operating characteristics, such as equal allocation of service level performance in excess of the associated goals, equal utilization of servers, damping of fluctuations in service level attainment, damping of fluctuations in server idleness, and mode-specific behavior. The composite preference evaluator 910 may be considered to constitute another layer of control on top of the previously described composite preference value function in this embodiment of the invention.

The composite preference evaluator 910 may dynamically select the composite preference function contingent upon various operating modes and various other operating conditions independent of normal service goals. Such modes might

be related to secondary goals involving efficiency of the server population, fairness considerations, job enrichment desires, or particular exceptional conditions.

Figure 9B is a flowchart illustrating some of the operations performed by the composite preference evaluator 910, according to an embodiment of the invention.

The composite preference evaluator 910 evaluates the enterprise and its associated work processing states to determine the overall attainment level of the enterprise's objectives for the work processing facility (step 901). The composite preference evaluator 910 then determines whether the presently selected user-selectable composite preference value function is the appropriate composite preference value function in light of the evaluation of the enterprise and its associated work processing states (step 902).

If the composite preference evaluator 910 determines that the presently selected user-selectable composite preference value function is the appropriate composite preference value function in light of the evaluation of the enterprise and its associated work processing states (step 902), then the composite preference evaluator 910 enters a wait state (step 903) before once again performing the enterprise evaluation (step 901).

If the composite preference evaluator 910 determines that the presently selected user-selectable composite preference value function is not the appropriate composite preference value function in light of the evaluation of the enterprise and its associated work processing states (step 902), then the composite preference evaluator 910 selects another user-selectable composite preference value function based on the present enterprise evaluation (step 904). The composite preference evaluator 910 then enters a wait state (step 903) before once again performing the enterprise evaluation (step 901).

The wait state (step 903) could terminate by a number of means and for a variety of reasons. For example, the wait state could be a predetermined period of time. Additionally, it could be triggered contingent upon or synchronous with the utilization of the composite preference value function or other events. As previously

discussed, the confluence of many types of attributes can determine the work types, according to an embodiment of the invention. The work type may be determined by multiple considerations, such as a combination of the task required, the product, the communications medium, geographic, governmental jurisdiction, demographic, language, skill, type of service, customer profile, stage in sales process, and regulatory classes, according to embodiments of the invention.

Figure 10 is a table 1000 illustrating server preferences and thresholds for work types performed in the work processing facility coupled with a composite preference value, according to an embodiment of the invention. The table 1000 generally resembles the table 300 shown in Figure 3 except that the work types in the work type column 302 may contain more than a single attribute. For example, a row 1001 contains a multiple attributes work type "calls from prospects in the western region." A corresponding row 306 in the table 300 contains the single attribute work type "calls from prospects." Assigning multiple attributes in a work type may increase the complexity of the over-all system and the total number of work type entries. While the table 1000 only lists the attributes for a server ID 100, similar information could be entered for other servers.

Consider the following example using the table 1000: an inquiry to a cellular telephone company may be classified geographically in order to satisfy different processing constraints in addition to classifying the inquiry as a work type for determination of server preference. Work items entering the work processing facility may not only be categorized for a first attribute (e.g., "prospect calls" or "claims calls") but may also be classified according to a second attribute (e.g., a regional description). Work items may also be classified according to various market segments to facilitate a market segmentation strategy. Demographic factors such as age, income, education, occupation, and household size could be used to identify customers in various target markets for various product lines. Typically, combinations of work and customer attributes effect a very complicated set of work types. The present invention supports such complications without the inefficiencies of unduly fragmenting the server population into small static pools.

The table 1000 also represents the need for some work types to have high priorities although the work type may not be performed frequently. For example, the work processing facility may operate under various regulatory constraints, such as the "public utility mandated calls" work type shown in a row 1005. The "public utility mandated calls" work type may comprise calls having very little profitability for the enterprise but calls that must nevertheless receive service at some officially mandated minimum/maximum level. Accordingly, the table 1000 illustrates a high composite preference value in the row 1005, indicating that the "public utility mandated calls" has come close to reaching (or even dropping below), the publicly mandated service level. Once the server assignor 102 has re-attained the required service level for the "public utility mandated calls," then the composite preference value drops from its high value. In addition to regulatory constraints, the table 1000 may contain other important, high-value goals that may be of limited importance to the enterprise otherwise, such as the attainment of service levels mandated by agreements with various telecommunications service workers unions.

As previously discussed, the server dispatch procedure may be both centralized and distributed. Figure 11 illustrates a distributed server assignment system 1101, according to an embodiment of the invention. The distributed server assignment system 1101 comprises computer systems 100a-100f. The computer systems 100a-100f contain server assignors 102a-102f and otherwise resemble to computer system 100 shown in Figure 1.

The server assignors 102a-102f may operate in the manner thus described for the server assignor 102. The server assignors 102a-102f may be located within a single work processing facility or spread across multiple work processing facilities that share work items (or pieces of work items) with each other.

When one of the server assignors 102a-102f cannot accomplish its server assignment tasks, then the server assignor communicates to the other server assignors using a network 1100. The server assignors 102a-102f may also communicate status information with each other using the network 1100.

Some of the server assignors 102a-102f may perform the dynamic preference value determination while other server assignors 102a-102f determine the composite preference value function, according to another embodiment of the invention. In this embodiment, there may be no communication between server assignors, and they act independently, except for communications with a specialized server assignor(s) that provides the dynamic preference value determination.

While this invention has been shown and described with references to preferred embodiments, it will be understood by those skilled in the art that various changes or modifications in form and detail may be made without departing from the scope of the invention. For example, the preference values utilized by the server assignor may be used to assign resources of all types. Further, preference values that are not contiguous integers may be assigned. Further, preference levels that are not contiguous integers may be assigned by the work processing facility, as may be non-numeric service levels. For example, the criteria for meeting good or bad service on a particular work type could avoid the use of integer or decimal numbers by using instead a Boolean function of transaction attributes or even some form of alphanumeric indicators, e.g., A1>C3. Moreover, a skilled artisan may readily utilize well-known empirical procedures for determining exemplary values to use for the server's preference values, threshold values, and dynamic preference values in any given work-processing facility.

While the description has discussed servers as belonging to server pools, a skilled artisan will recognize that the servers need never really be identified as members of a particular server pool. In other words, the servers may merely receive work of one type or another, hence they may be thought of as belonging to a server pool for that work type even though the pool may not physically exist.

The invention is also applicable to systems in which work items are described with regard to both a work type and a service class. The service class may describe particular aspects of the work item while the work type describes the work item from the server's point of view. Accordingly, in some embodiment of the

invention, the service class may influence the dynamic preference value and ultimately the composite preference value.

While the examples discussed above relate to calls in a call center, the invention is not limited to the processing of calls in a call center by servers, both human and/or robotic. The invention may be applied to the servicing of any form of work type and any form of service class. For example, the invention is equally applicable to the processing of electronic mail messages and video streaming tasks.

These and other changes can be made to the invention in light of the above detailed description. In general, in the following claims, the terms used should not be construed to limit the invention to the specific embodiments disclosed in the specification and the claims, but should be construed to include all hierarchical remediation methods and systems that operate in accordance with the invention. Accordingly, the invention is not limited by the disclosure, but instead its scope is to be determined by the following claims.

4. Brief Description of Drawings

An embodiment of the invention will be described below relative to the following figures. Note that similar elements and steps in the figures have the same reference number.

Figure 1 is a high-level block diagram of a sample general-purpose computer system 100 upon which a server assignor 102 may execute, according to an embodiment of the invention.

Figure 2 is a server preference table 200 illustrating the preferences of various servers for the various work types 202 that represent types of work performed in the work processing facility, according to an embodiment of the invention.

Figure 3 is a diagram showing a table 300 illustrating server preferences and thresholds for work types performed in the work processing facility coupled with a composite preference value for the work type, according to an embodiment of the invention.

Figure 4 is a flowchart illustrating the generation of composite preference values using the server assignor 102 or its corresponding method, according to an embodiment of the invention.

Figures 5A and 5B are a flowchart illustrating the server dispatch procedure, according to an embodiment of the invention.

Figures 6A-6D illustrate the change in assignment of servers to various server pools 603-605 over time due to the server assignor 102, according to an embodiment of the invention.

Figure 7 is a diagram showing a table 700 that includes a dynamic preference value column 701, according to an alternative embodiment of the invention.

Figure 8 is a server preference table 800 illustrating the preferences of various servers for the various work types 202 that represent types of work performed in the work processing facility, according to an embodiment of the invention.

Figure 9A is a high-level block diagram of another general-purpose computer system 100 that includes a threshold value determiner 911 and a composite preference evaluator 910, according to an alternative embodiment of the invention.

Figure 9B is a flowchart illustrating some of the operations performed by the composite preference evaluator 910, according to an embodiment of the invention.

Figure 10 is a table 1000 illustrating server preferences and thresholds for work types performed in another work processing facility coupled with a composite preference value, according to an embodiment of the invention.

Figure 11 illustrates a distributed server assignment system 1101, according to an embodiment of the invention.

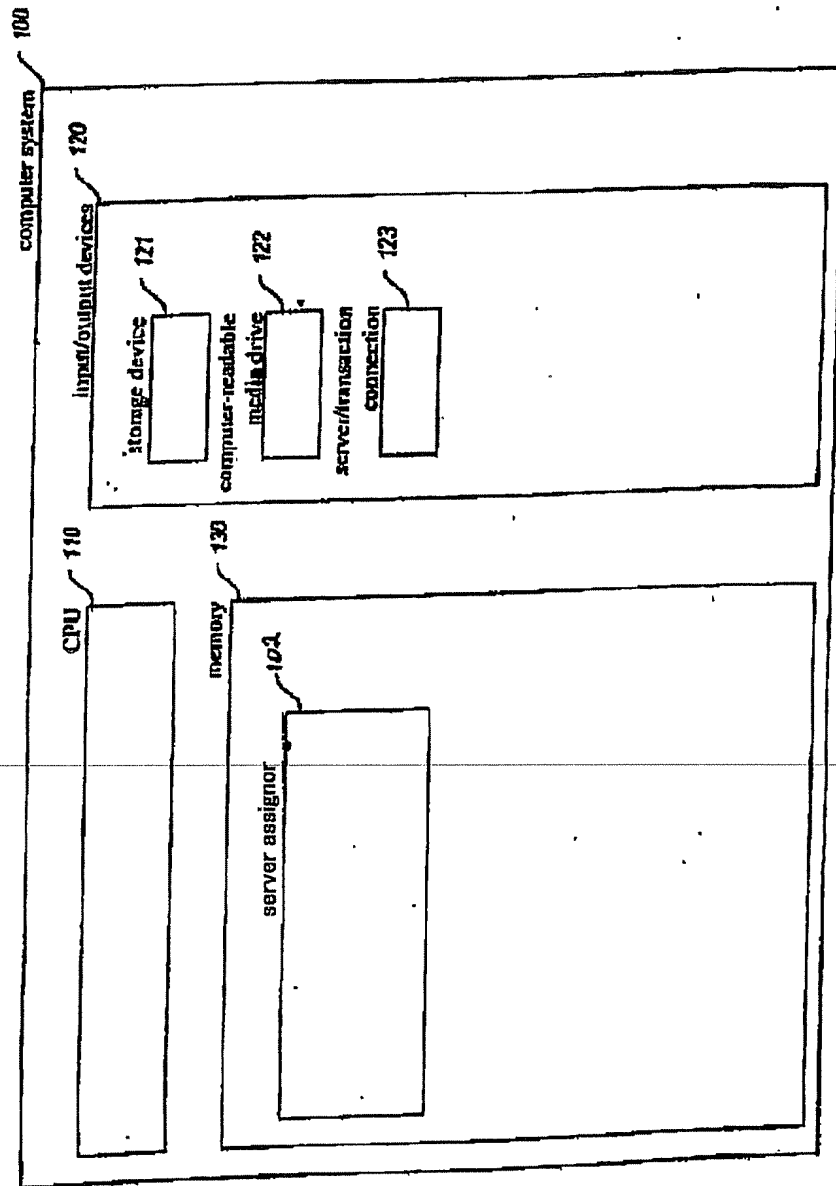


Fig. 1

Server ID	Work Type	Server Preference for Work Type	Server Threshold for Work Type
100	Calls from Prospects	8	9
100	Claims Calls	4	3
101	Calls from Prospects	5	9
101	Claims Calls	6	7
101	Calls from Preferred Customers	0	0

Fig. 2

Server ID	Work Type	Server Preference for Work Type	Server Threshold for Work Type	Composite Preference for Work Type
100	Calls from Prospects	8	9	10
100	Claims Calls	4	3	8
101	Calls from Prospects	5	9	7
101	Claims Calls	6	7	10
101	Calls from Preferred Customers	0	0	0

Fig. 3

Fig.4

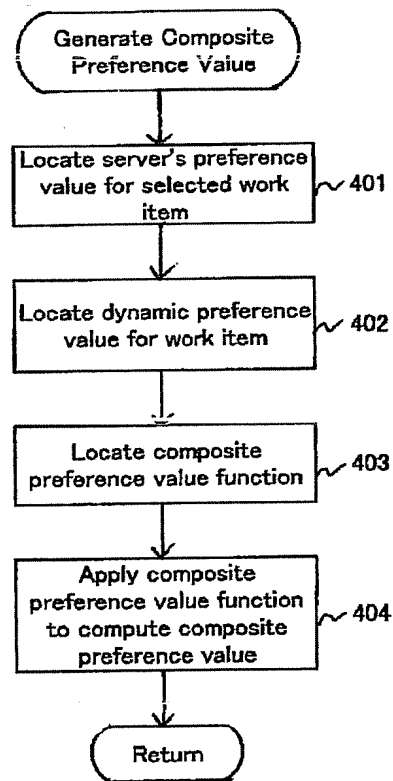


Fig.5A

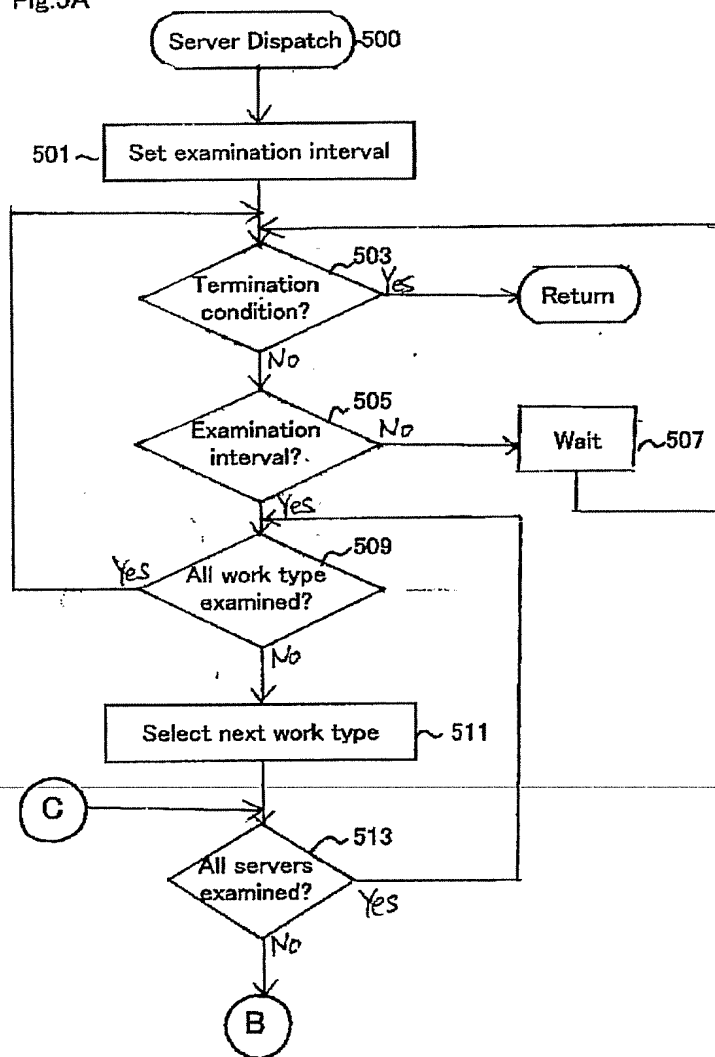


Fig.5B

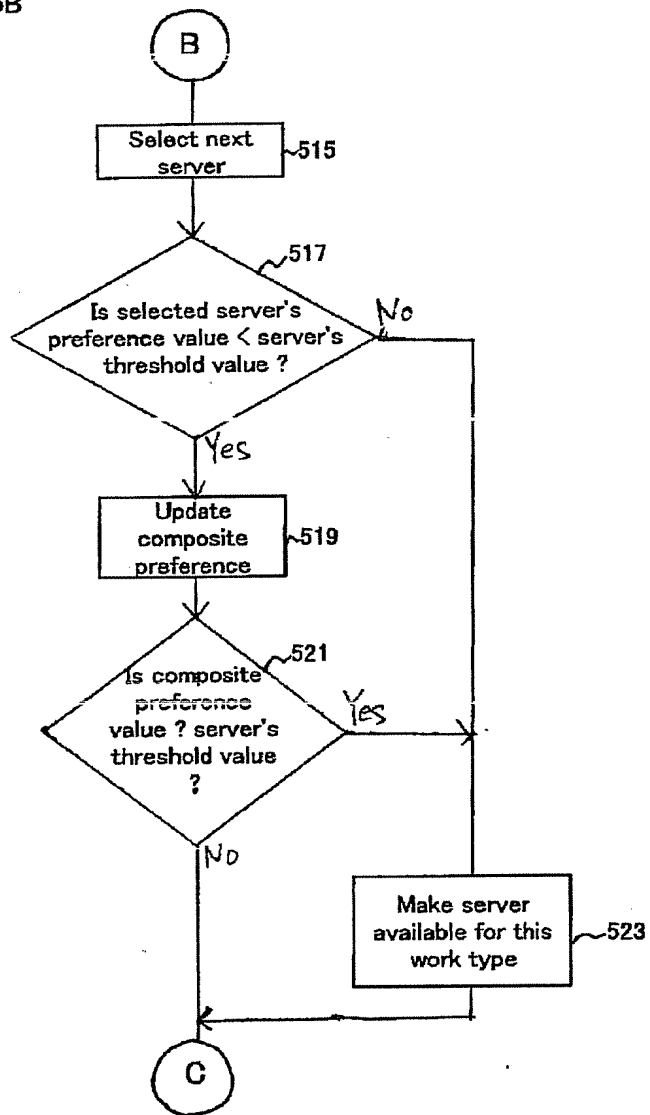


Fig.6A

Work Type	Available Servers	
Calls from Prospects	42	603
Claims Calls	30	604
Calls from Preferred Customers	15	605

Fig.6B

Work Type	Available Servers	
Calls from Prospects	48	603
Claims Calls	24	604
Calls from Preferred Customers	15	605

Fig.6C

Work Type	Available Servers	
Calls from Prospects	53	603
Claims Calls	24	604
Calls from Preferred Customers	10	605

Fig.6D

Work Type	Available Servers	
Calls from Prospects	50	603
Claims Calls	26	604
Calls from Preferred Customers	11	605

Server ID	Work Type	Server Preference for Work Type	Server Threshold for Work Type	Dynamic Preference Value for Work Type	Composite Preference for Work Type
100	Calls from Prospects	8	9	2	10
100	Claims Calls	4	3	4	8
101	Calls from Prospects	5	9	2	7
101	Claims Calls	6	7	4	10
101	Calls from Preferred Customers	0	0	0	0

Fig. 7

Fig.8

201 Server ID	202 Work Type	203 Server Preference for Work Type
100	Calls from Prospects	8
100	Claims Calls	4
101	Calls from Prospects	5
101	Claims Calls	6
101	Calls from Preferred Customers	0

806
Threshold Value Table

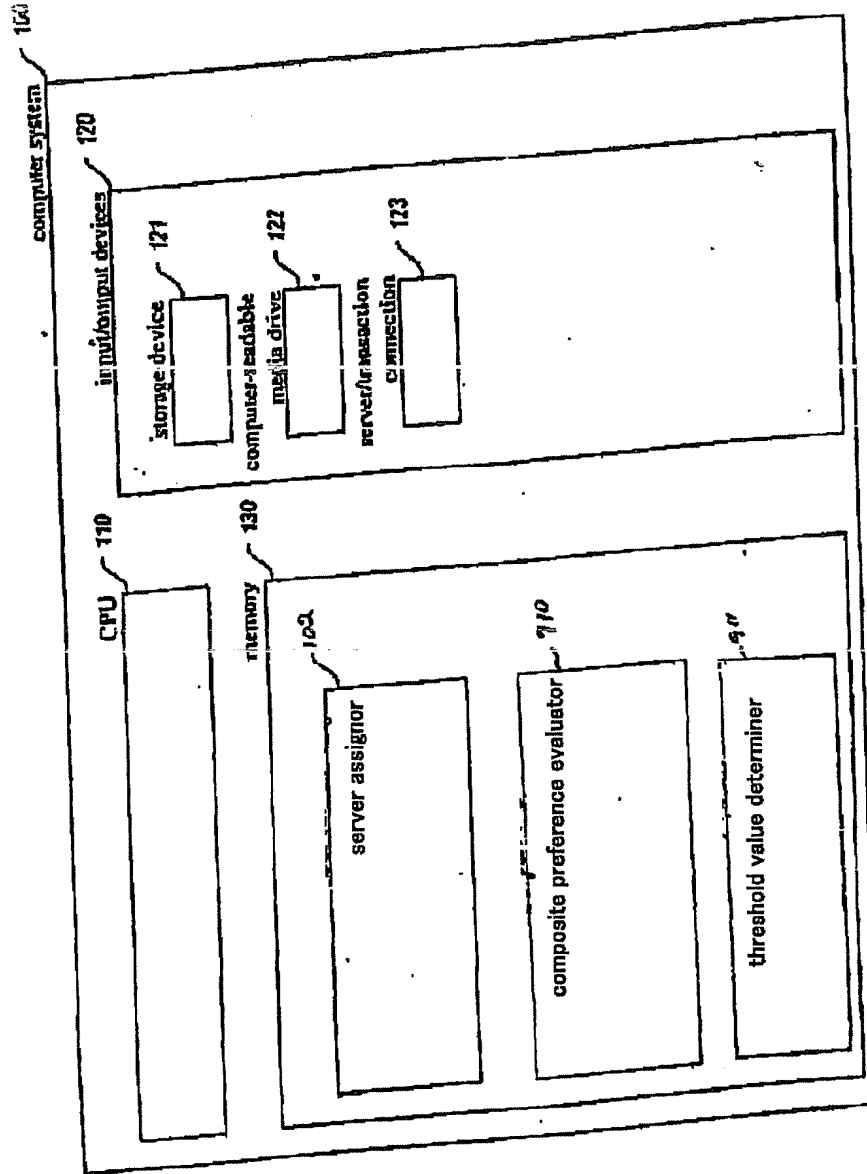


Fig. 9A

Fig.9B

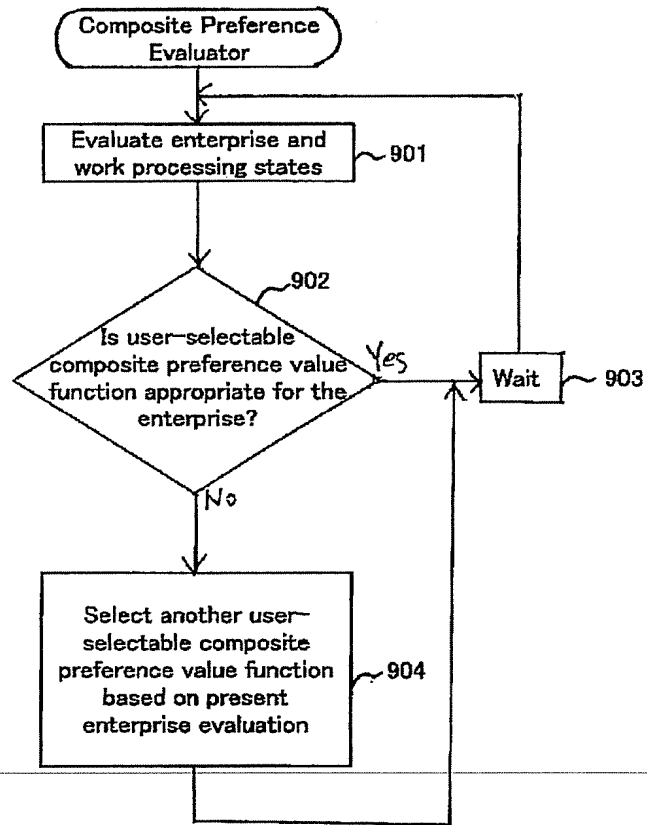
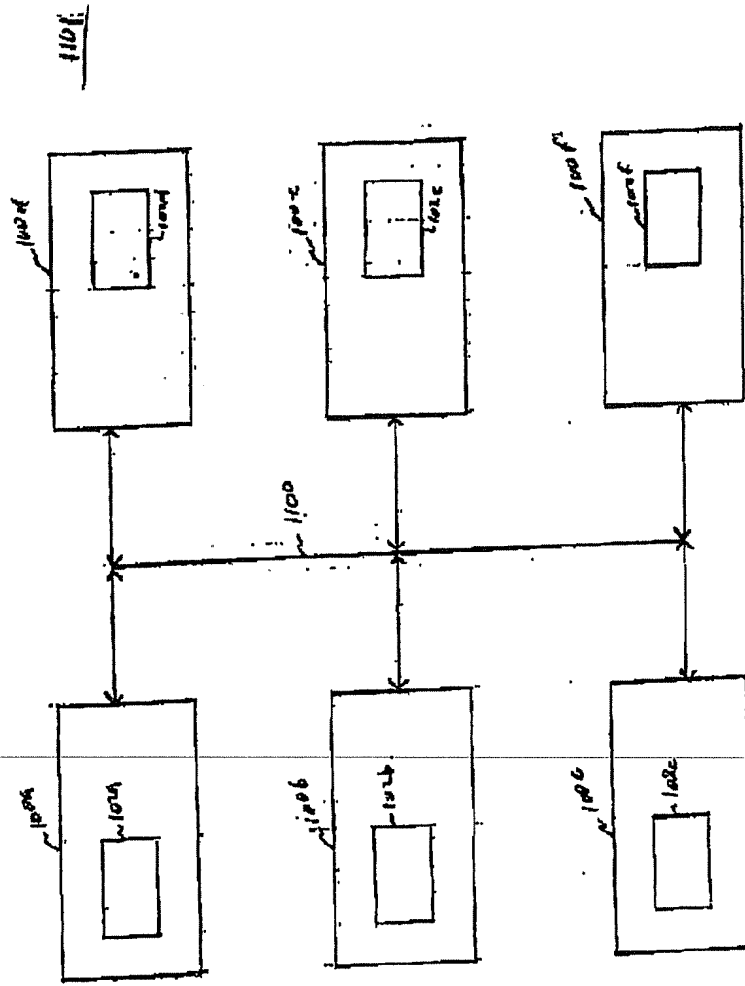


FIG.10

Agent ID	Work Type	Server Preference for Work Type	Server Threshold for Work Type	Composite Preference for Work Type
100	Calls from Prospects in Western Region	8	9	10
100	Claims Calls from Western Region	4	3	8
100	Calls from Prospects in Central Region	5	9	7
100	Claims Calls from Central Region	6	7	10
100	Public Utility Mandated Calls	8	9	20



Abstract

A facility for adjusting a number of servers assigned to server pools for performing certain work types on the basis of unmet service needs in a work processing facility. Servers may include service agents, both human and robotic. A server assignor and a corresponding server assignment method may each be employed in a work distributor or an automatic call distributor ("ACD") to conditionally adjust server availability in server pools. The server assignor compares a composite preference value for a work type against each server's threshold value for the work type. When the server assignor determines that the composite preference value is greater than or equal to a server's threshold value, then the server assignor indicates that the server may be included in the server pool for that work type. Each server has preference values and threshold values for different kinds of work. The magnitude of a preference value represents an affinity for the work type. The server's threshold value represents a reluctance to perform work having that work type. The server does not normally receive work for which the preference value is less than the threshold value, unless the server assignor determines that the composite preference value exceeds the server's threshold value for that work type. The server assignor and corresponding server assignment method may compute the composite preference value from a number of user-selectable inputs and utilize a number of user-selectable functions.

2 Representative Drawing

Figure 1